# Contents

# Chapter 1

# Preamble

## 1.1 Additional Suggested Reading

In addition to the supplementary readings mentioned in the endnotes, there are three larger collections of articles pertinent to this Preamble. We give the titles, sources, and general classification headings below:

A.1 – General Ethical Issues

A.1.1 – Editorial: Chance, or Human Judgment? (Dael Wolfle, *Science*, February 27, 1970)

A.1.2 – Wall Street Smarts (Calvin Trillin, *New York Times*, October 14, 2009)

A.1.3 – Teaching Physicians the Price of Care (Susan Okie, *New York Times*, May 3, 2010)

A.1.4 – Why I Wrote "The Crucible" (Arthur Miller, *New Yorker*, October 10, 1996)

A.1.5 – What Every Girl Should Know (Gail Collins, *New York Times*, May 8, 2010)

A.1.6 – Researchers Say They Created a 'Synthetic Cell' (Nicholas Wade, *New York Times*, May 20, 2010)

A.1.7 – New York Minorities More Likely to Be Frisked (Al Baker, *New York Times*, May 12, 2010)

A.2 – Clinical Psychology and Psychiatry

A.2.1 – Defying Psychiatric Wisdom, These Skeptics Say 'Prove It' (Erica Goode, *New York Times*, March 9, 2004)

A.2.2 – Editorial — Psychology: A Reality Check (*Nature*, October 15, 2009)

A.2.3 – Ignoring the Evidence: Why Do Psychologists Reject Science? (Sharon Begley, *Newsweek*, October 2, 2009)

A.2.4 – The Trauma Trap (Frederick C. Crews, *New York Review of Books*, October 14, 2009)

A.3 – Ethically Questionable Appeals

A.3.1 – Sure, It's Treatable. But Is It a Disorder? (Natasha Singer, *New York Times*, December 13, 2009)

A.3.2 – Menopause, as Brought to You by Big Pharma (Natasha Singer and Duff Wilson, *New York Times*, December 13, 2009)

A.3.3 – Cancer Center Ads Use Emotion More Than Fact (Natasha Singer, *New York Times*, December 19, 2009)

A.3.4 – Ask Your Doctor If Direct-to-Consumer Healthcare Advertising Is Right for You (*Knowledge@W.P. Carey*, February 13, 2008)

A.3.5 – Flavored Tobacco Pellets Are Denounced as a Lure to Young Users (Duff Wilson, *New York Times*, April 18, 2010)

A.3.6 – The Hard Sell on Salt (Michael Moss, *New York Times*, May 29, 2010)

A.3.7 – The Menopausal Marketplace (Amanda Spake, *US News and World Report*, November 10, 2001)

A.3.8 – New Blood-Pressure Guidelines Pay Off – For Drug Companies (Duff Wilson, *The Seattle Times*, June 26, 2005)

A.3.9 – Rush Toward New Weight-Loss Drugs Tramples Patients' Health (Susan Kelleher, *The Seattle Times*, June 27, 2005)

A.3.10 – Disease Expands Through Marriage of Marketing and Machines (Susan Kelleher, *The Seattle Times*, June 28, 2005)

A.3.11 – Many New Drugs Have Strong Dose of Media Hype (Duff Wilson, *The Seattle Times*, June 29, 2005)

A.3.12 – Clash Over "Little Blue Pill" for Women (Susan Kelleher, *The Seattle Times*, June 30, 2005)

A.4 – Publication

A.4.1 – Medical Papers by Ghostwriters Pushed Therapy (Natasha Singer, *New York Times*, August 5, 2009)

A.4.2 – Integrity Under Attack: The State of Scholarly Publishing (Douglas N. Arnold, *SIAM News*, December, 2009)

A.4.3 – Influence of Funding Source on Outcome, Validity, and Reliability of Pharmaceutical Research (Council on Scientific Affairs of the American Medical Association)

A.4.4 – Sponsorship, Authorship, and Accountability: International Committee of Medical Journal Editors (August, 2007)

A.4.5 – Journal Retracts 1998 Paper Linking Autism to Vaccines (Gardiner Harris, *New York Times*, February 3, 2010)

A.4.6 – Journal Retracts Flawed Study Linking MMR Vaccine and Autism (Nathan Seppa, *Science News*, February 3, 2010)

A.4.7 – Stop Misbehaving! (Ushma S. Neill, *The Journal of Clinical Investigation*, July 3, 2006)

A.4.8 – The Politics of Sex Abuse (Carol Tavris, *The Los Angeles Times*, July 19, 1999)

A.4.9 – A Funny Thing Happened on the Way to My *American Psychologist* Publication (Scott O. Lilienfeld, *American Psychologist*, March, 2002)

A.4.10 – Long After Kinsey, Only the Brave Study Sex (Benedict Cary, *New York Times*, November 9, 2004)

A.4.11 – Study Sees a Slant in Articles on Drug (Nicholas Bakalar, *New York Times*, April 12, 2010)

A.4.12 – Maternal Deaths Decline Sharply Across the Globe (Denise Grady, *New York Times*, April 13, 2010)

A.4.13 – Atrazine Paper's Challenge: Who's Responsible for Accuracy? (Janet Raloff, *Science News*, May 6, 2010)

A.4.14 – British Medical Council Bars Doctor Who Linked Vaccine With Autism (John G. Burns, *New York Times*, May 24, 2010)

A.4.15 – Citation Inflation (Janet Raloff, *Science News*, June 15, 2010)

A.4.16 – Report Urges More Curb on Medical Ghostwriting (Natasha Singer, *New York Times*, June 24, 2010)

A.4.17 – Scientists Criticize Study on Genetics of Old Age (Nicholas Wade, *New York Times*, July 8, 2010)

A.5 – Questionable Practice: Radiation

A.5.1 – Side Effects: Profit and Questions on Prostate Cancer Therapy (Stephanie Saul, *New York Times*, December 1, 2006)

A.5.2 – Sales Pitch for a Treatment (Stephanie Saul, *New York Times*, December 1, 2006)

A.5.3 – Hospitals Look to Nuclear Tool to Fight Cancer (Andrew Pollack, *New York Times*, December 26, 2007)

A.5.4 – The Radiation Boom: Radiation Offers New Cures, and Ways to Do Harm (Walt Bogdanich, *New York Times*, January 24, 2010)

A.5.5 – The Radiation Boom: As Technology Surges, Radiation Safeguards Lag (Walt Bogdanich, *New York Times*, January 27, 2010)

A.5.6 – Medical Group Urges New Rules on Radiation (Walt Bogdanich, *New York Times*, February 5, 2010)

A.5.7 – F.D.A. to Increase Oversight of Medical Radiation (Walt Bogdanich and Rebecca R. Ruiz, *New York Times*, February 10, 2010)

A.5.8 – Radiation Errors Reported in Missouri (Walt Bogdanich and Rebecca R. Ruiz, *New York Times*, February 24, 2010)

A.5.9 – Radiation Bills Raise Question of Supervision (Walt Bogdanich and Rebecca R. Ruiz, *New York Times*, February 25, 2010)

A.5.10 – At Hearing on Radiation, Calls for Better Oversight (Walt Bogdanich, *New York Times*, February 27, 2010)

A.5.11 – Scientists Say F.D.A. Ignored Radiation Warnings (Gardiner Harris, *New York Times*, March 28, 2010)

A.5.12 – F.D.A. Toughens Process for Radiation Equipment (Walt Bogdanich, *New York Times*, April 8, 2010)

A.5.13 – Americans Get Most Radiation From Medical Scans (Marilynn Marchione, *The Associated Press*, July 14, 2010)

A.6 – Questionable Practice: Big Pharma

A.6.1 – Drug Companies & Doctors: A Story of Corruption (Marcia Angell, *New York Review of Books*, January 15, 2009)

A Note to Readers: By The Editors of the *New York Review of Books*

'Drug Companies & Doctors': An Exchange

A.6.2 – How a New Policy Led to Seven Deadly Drugs (David Willman, *Los Angeles Times*, December 20, 2000)

A.6.3 – Dangerous Data: Despite Warnings, Drug Giant Took Long Path to Vioxx Recall (Alex Berenson et al., *New York Times*, November 14, 2004)

A.6.4 – Research Ties Diabetes Drug to Heart Woes (Gardiner Harris, *New York Times*, February 20, 2010)

A.6.5 – A Face-Off on the Safety of a Drug for Diabetes (Gardiner Harris, *New York Times*, February 23, 2010)

A.6.6 – Biologics Boondoggle (Anthony D. So and Samuel L. Katz, *New York Times*, March 8, 2010)

A.6.7 – Sought-After Speaker, With Script Outlines From Eli Lilly (Natasha Singer, *New York Times*, November 4, 2009)

A.6.8 – Stanford Medical School to Expand Ethics Rules (Natasha Singer, *New York Times*, March 21, 2010)

A.6.9 – Eli Lilly Said to Play Down Risk of Top Pill (Alex Berenson, *New York Times*, December 17, 2006)

A.6.10 – Judge to Unseal Documents on the Eli Lilly Drug Zyprexa (Mary Williams Walsh, *New York Times*, September 6, 2008)

A.6.11 – When Heart Devices Fail, Who Should Be Blamed? (Barry Meier, *New York Times*, April 20, 2010)

A.6.12 – F.D.A. Again Warns a Generic Maker About Conditions at Its Plants (Natasha Singer, *New York Times*, April 15, 2010)

A.6.13 – Tylenol, Generics and Trust (Natasha Singer, *New York Times*, May 2, 2010)

A.6.14 – Agency Told Tylenol Maker of Many Quality Concerns (Natasha Singer, *New York Times*, May 4, 2010)

A.6.15 – Data on Fees to Doctors Is Called Hard to Parse (Duff Wilson, *New York Times*, April 12, 2010)

A.6.16 – Safety Rules Can't Keep Up With Biotech Industry (Andrew Pollack and Duff Wilson, *New York Times*, May 27, 2010)

A.6.17 – Questions for Makers on Defects in Drugs (Natasha Singer, *New York Times*, May 26, 2010)

A.6.17 – F.D.A. Weighs Penalties in Drug Recall (Natasha Singer, *New York Times*, May 27, 2010)

A.6.18 – Caustic Government Report Deals Blow to Diabetes Drug (Gardiner Harris, *New York Times*, July 9, 2010)

A.6.20 – Diabetes Drug Maker Hid Test Data on Risks, Files Indication (Gardiner Harris, *New York Times*, July 12, 2010)

# Chapter 2

# Introduction

# Chapter 3

# Probability Theory: Background and Bayes Theorem

## 3.1 Suggested Reading Relevant to the Whole Chapter

B.1 – Generally Relevant for This Chapter

B.1.1 – The Prosecutor's Fallacy (Mark Buchanan, *New York Times*, May 16, 2007)

B.1.2 – Better Decisions Through Science (John A. Swets, Robyn M. Dawes, and John Monahan, *Scientific American*, October 2000)

B.1.3 – Maybe We Should Leave That Up to the Computer (Douglas Meingartner, *New York Times*, July 18, 2006)

## 3.2 Suggested Reading on Probability Issues

B.2 – Probability Issues

B.2.1 – Trawling the Brain (Laura Sanders, *Science News*, October 19, 2009)

B.2.2 – The Cancer-Cluster Myth (Atul Gawande, *New Yorker*, February 8, 1999)

B.2.3 – If Smallpox Strikes Portland ... (Chris L. Barrett, Stephen G. Eubank, and James P. Smith, *Scientific American*, February 21, 2005)

B.2.4 – The Treatment (Malcolm Gladwell, *New Yorker*, May 17, 2010)

B.2.5 – All Present-Day Life Arose From a Single Ancestor (Tina Hesman Saey, *Science News*, June 5, 2010)

B.2.6 – Nabbing Suspicious SNPS: Scientists Search the Whole Genome for Clues to Common Diseases (Regina Nuzzo, *Science News*, June 21, 2008)

## 3.3 Suggested Reading on Forensic Issues

B.3 – Forensic Issues

B.3.1 – Do Fingerprints Lie? (Michael Spector, *New Yorker*, May 27, 2002)

B.3.2 – You Think DNA Evidence Is Foolproof? Try Again (Adam Liptak, *New York Times*, March 16, 2003)

B.3.3 – Under Suspicion (Atul Gawande, *New Yorker*, January 8, 2001)

B.3.4 – Line-Ups On Trial (Laura Spinney, *Nature*, May 22, 2008)

B.3.5 – Florida Man Exonerated After 35 Years Behind Bars (Mitch Stacy, *Associated Press*, December 17, 2009)

B.3.6 – How Police Interrogation Works (Julia Layton)

B.3.7 – Duped (Margaret Talbot, *New Yorker*, July 2, 2007)

B.3.8 – The Truth Hurts (Rachel Ehrenberg, *Science News*, July 3, 2010)

## 3.4 Suggested Reading on Screening

B.4 – Screening

B.4.1 – In Shift, Cancer Society Has Concerns on Screenings (Gina Kolata, *New York Times*, October 21, 2009)

B.4.2 – Addicted to Mammograms (Robert Aronowitz, *New York Times*, November 20, 2009)

B.4.3 – Screening Debate Reveals Culture Clash in Medicine (Kevin Sack, *New York Times*), November 20, 2009)

B.4.4 – The Breast Brouhaha (Gail Collins, *New York Times*, November 19, 2009)

B.4.5 – Behind Cancer Guidelines, Quest for Data (Gina Kolata, *New York Times*, November 23, 2009)

B.4.6 – Mammogram Math (John Allen Paulos, *New York Times*, December 13, 2009)

B.4.7 – Gauging the Odds (and the Costs) in Health Screening (Richard H. Thaler, *New York Times*), December 20, 2009)

B.4.8 – The Risks and Benefits of Cancer Screenings (Sandra G. Boodman, *AARP Bulletin*, January 1, 2010)

B.4.9 – Chances Are (Steven Strogatz, *New York Times*, April 25, 2010)

B.4.10 – Cancer Fight: Unclear Tests for New Drug (Gina Kolata, *New York Times*, April 19, 2010)

B.4.11 – Blood Test for Early Ovarian Cancer May Be Recommended for All (Tom Randall, *Bloomberg*, May 21, 2010)

B.4.12 – Doubt Is Cast on Many Reports of Food Allergies (Gina Kolata, *New York Times*, May 11, 2010)

B.4.13 – I Can't Eat That. I'm Allergic (Gina Kolata, *New York Times*, May 14, 2010)

B.4.14 – Promise Seen for Detection of Alzheimer's (Gina Kolata, *New York Times*, June 23, 2010)

B.4.15 – Breast Screening Tool Finds Many Missed Cancers (Janet Raloff, *Science News*, July 1, 2010)

# Chapter 4

# Probability Theory: Application Areas

## 4.1 Suggested Reading on Agent Orange and Judge Weinstein

In the Suggested Reading of Section B.5 (Agent Orange and Judge Weinstein), we list seven items. Five deal with the path that the Agent Orange litigation took in Judge Weinstein's court since the early 1980s:

B.5 – Agent Orange and Judge Weinstein

B.5.1 – Vietnam Agent Orange Suit By Veterans Is Going to Trial (Ralph Blumenthal, *New York Times*, May 6, 1984)

B.5.2 – Lack of Military Data Halts Agent Orange Study (Philip M. Boffey, *New York Times*, September 1, 1987)

B.5.3 – Agent Orange, the Next Generation; In Vietnam and America, Some See a Wrong Still Not Righted (William Glaberson, *New York Times*, August 8, 2004)

B.5.4 – Brooklyn: Agent Orange Lawsuits Dismissed (William Glaberson, *New York Times*, November 19, 2004)

B.5.5 – Civil Lawsuit on Defoliant in Vietnam Is Dismissed (William Glaberson, *New York Times*, March 11, 2005)

B.5.6 – High Caliber Justice (Robert Kolker, *New York Magazine*, April 5, 1999)

B.5.7 – Defiant Judge Takes On Child Pornography Law (A. G. Sulzberger, *New York Times*, May 21, 2010)

The penultimate piece listed as B.5.6 is a profile of Judge Weinstein, *High Caliber Justice*, from *New York Magazine* (Robert Kolker, April 5, 1999). The final item (B.5.7) regards a recent kerfuffle that Judge Weinstein raised with respect to mandatory sentences for the possession of child pornography (A. G. Sulzberger, *Defiant Judge Takes On Child Pornography Law*, *New York Times*, May 21, 2010). The issue concerns what a judge can do when disagreeing with an imposed mapping (created by legislative action) between "the crime"

and "the amount of time to serve." In the past, Judge Weinstein has refused to preside over many drug trials when he felt his hands would be tied inappropriately regarding final sentencing.

### 4.1.1 Appendix: The Redacted Text of Judge Weinstein's Opinion in the Fatico Case

We begin with the caution of Justice Brennan in Speiser v. Randall, about the crucial nature of fact finding procedures:

To experienced lawyers it is commonplace that the outcome of a lawsuit and hence the vindication of legal rights depends more often on how the factfinder appraises the facts than on a disputed construction of a statute or interpretation of a line of precedents. Thus the procedures by which the facts of the case are determined assume an importance fully as great as the validity of the substantive rule of law to be applied. And the more important the rights at stake, the more important must be the procedural safeguards surrounding those rights.

The "question of what degree of proof is required ... is the kind of question which has traditionally been left to the judiciary to resolve ... "

Broadly stated, the standard of proof reflects the risk of winning or losing a given adversary proceeding or, stated differently, the certainty with which the party bearing the burden of proof must convince the factfinder.

As Justice Harlan explained in his concurrence in Winship, the choice of an appropriate burden of proof depends in large measure on society's assessment of the stakes involved in a judicial proceeding.

In a judicial proceeding in which there is a dispute about the facts of some earlier event, the factfinder cannot acquire unassailably accurate knowledge of what happened. Instead, all the factfinder can acquire is a belief of what Probably happened. The intensity of this belief—the degree to which a factfinder is convinced that a given act actually occurred—can, of course, vary. In this regard, a standard of proof represents an attempt to instruct the factfinder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for a particular type of adjudication. Although the phrases "preponderance of the evidence" and "proof beyond a reasonable doubt" are quantitatively imprecise, they do communicate to the finder of fact different notions concerning the degree of confidence he is expected to have in the correctness of his factual conclusions.

Thus, the burden of proof in any particular class of cases lies along a continuum from low probability to very high probability.

Preponderance of the Evidence:

As a general rule, a "preponderance of the evidence" [or] more probable than not standard, is relied upon in civil suits where the law is indifferent as between plaintiffs and defendants, but seeks to minimize the probability of error.

In a civil suit between two private parties for money damages, for example, we view it as no more serious in general for there to be an erroneous verdict in the defendant's

favor than for there to be an erroneous verdict in the plaintiff's favor. A preponderance of the evidence standard therefore seems peculiarly appropriate; as explained most sensibly, it simply requires the trier of fact "to believe that the existence of a fact is more probable than its nonexistence before (he) may find in favor of the party who has the burden to persuade the (judge) of the fact's existence."

Quantified, the preponderance standard would be 50+% Probable.

Clear and Convincing Evidence:

In some civil proceedings where moral turpitude is implied, the courts utilize the standard of "clear and convincing evidence," a test somewhat stricter than preponderance of the evidence.

Where proof of another crime is being used as relevant evidence pursuant to Rules 401 to 404 of the Federal Rules of Evidence, the most common test articulated is some form of the "clear and convincing" standard.

Quantified, the probabilities might be in the order of above 70% under a clear and convincing evidence burden.

Clear, Unequivocal and Convincing Evidence:

"In situations where the various interests of society are pitted against restrictions on the liberty of the individual, a more demanding standard is frequently imposed, such as proof by clear, unequivocal and convincing evidence." The Supreme Court has applied this stricter standard to deportation proceedings, denaturalization cases, and expatriation cases. In Woodby, the Court explained:

To be sure, a deportation proceeding is not a criminal prosecution. But it does not syllogistically follow that a person may be banished from this country upon no higher degree of proof than applies in a negligence case. This Court has not closed its eyes to the drastic deprivations that may follow when a resident of this country is compelled by our Government to forsake all the bonds formed here and go to a foreign land where he often has no contemporary identification.

In terms of percentages, the probabilities for clear, unequivocal and convincing evidence might be in the order of above 80% under this standard.

Proof Beyond a Reasonable Doubt:

The standard of "proof beyond a reasonable doubt" is constitutionally mandated for elements of a criminal offense. Writing for the majority in Winship, Justice Brennan enumerated the "cogent reasons" why the "'reasonable-doubt' standard plays a vital role in the American scheme of criminal procedure" and "is a prime instrument for reducing the risk of convictions resting on factual error."

The accused during a criminal prosecution has at stake interest of immense importance, both because of the possibility that he may lose his liberty upon conviction and because of the certainty that he would be stigmatized by the conviction. Accordingly, a society that values the good name and freedom of every individual should not condemn a man for commission of a crime when there is reasonable doubt about his guilt. As we said in Speiser v. Randall, "There is always in litigation a margin of error, representing error in fact finding, which both parties must take into account. Where one party has at stake an interest of transcending

value as a criminal defendant—his liberty—this margin of error is reduced as to him by the process of placing on the other party the burden of ... persuading the factfinder at the conclusion of the trial of his guilt beyond a reasonable doubt. Due process commands that no man shall lose his liberty unless the Government has borne the burden of ... convincing the factfinder of his guilt." ...

Moreover, use of the reasonable-doubt standard is indispensable to command the respect and confidence of the community in applications of the criminal law. It is critical that the moral force of the criminal law not be diluted by a standard of proof that leaves people in doubt whether innocent men are being condemned.

In capital cases, the beyond a reasonable doubt standard has been utilized for findings of fact necessary to impose the death penalty after a finding of guilt.

Many state courts, in interpreting state recidivism statutes, have held that proof of past crimes must be established beyond a reasonable doubt.

In civil commitment cases, where the stakes most resemble those at risk in a criminal trial, some courts have held that the beyond a reasonable doubt standard is required.

If quantified, the beyond a reasonable doubt standard might be in the range of 95+% Probable.

### 4.1.2 Appendix: *Guidelines for Determining the Probability of Causation and Methods for Radiation Dose Reconstruction Under the Employees Occupational Illness Compensation Program Act of 2000*

SUMMARY: This rule implements select provisions of the Energy Employees Occupational Illness Compensation Program Act of 2000 ("EEOICPA" or "Act"). The Act requires the promulgation of guidelines, in the form of regulations, for determining whether an individual with cancer shall be found, "at least as likely as not," to have sustained that cancer from exposure to ionizing radiation in the performance of duty for nuclear weapons production programs of the Department of Energy and its predecessor agencies. The guidelines will be applied by the U.S. Department of Labor, which is responsible for determining whether to award compensation to individuals seeking federal compensation under the Act. (p. 22296)

B. Purpose of Probability of Causation Guidelines

Under EEOICPA, a covered employee seeking compensation for cancer, other than as a member of the Special Exposure Cohort seeking compensation for a specified cancer, is eligible for compensation only if DOL determines that the cancer was "at least as likely as not" (a 50% or greater probability) caused by radiation doses incurred in the performance of duty while working for DOE and/or an atomic weapons employer (AWE) facility. These guidelines provide DOL with the procedure to make these determinations, and specify the information DOL will use. (p. 22296)

D. Understanding Probability of Causation

Probability of Causation is a technical term generally meaning an estimate of the per-

centage of cases of illness caused by a health hazard among a group of persons exposed to the hazard. This estimate is used in compensation programs as an estimate of the probability or likelihood that the illness of an individual member of that group was caused by exposure to the health hazard. Other terms for this concept include "assigned share" and "attributable risk percent".

In this rule, the potential hazard is ionizing radiation to which U.S. nuclear weapons workers were exposed in the performance of duty; the illnesses are specific types of cancer. The probability of causation (PC) is calculated as the risk of cancer attributable to radiation exposure (RadRisk) divided by the sum of the baseline risk of cancer to the general population (BasRisk) plus the risk attributable to the radiation exposure, then multiplied by 100 percent, as follows:

$$\text{PC} = \frac{\text{RadRisk}}{\text{RadRisk} + \text{BasRisk}} \times 100\%$$

This calculation provides a percentage estimate between 0 and 100 percent, where 0 would mean 0 likelihood that radiation caused the cancer and 100 would mean 100 percent certainty that radiation caused the cancer.[1]

Scientists evaluate the likelihood that radiation caused cancer in a worker by using medical and scientific knowledge about the relationship between specific types and levels of radiation dose and the frequency of cancers in exposed populations. Simply explained, if research determines that a specific type of cancer occurs more frequently among a population exposed to a higher level of radiation than a comparable population (a population with less radiation exposure but similar in age, gender, and other factors that have a role in health), and if the radiation exposure levels are known in the two populations, then it is possible to estimate the proportion of cancers in the exposed population that may have been caused by a given level of radiation.

If scientists consider this research sufficient and of reasonable quality, they can then translate the findings into a series of mathematical equations that estimate how much the risk of cancer in a population would increase as the dose of radiation incurred by that population increases. The series of equations, known as a dose-response or quantitative risk assessment model, may also take into account other health factors potentially related to cancer risk, such as gender, smoking history, age at exposure (to radiation), and time since exposure. The risk models can then be applied as an imperfect but reasonable approach to determine the likelihood that the cancer of an individual worker was caused by his or her radiation dose.

E. Development and Use of the RadioEpidemiological Tables and Interactive RadioEpidemiological Program

In 1985, in response to a congressional mandate in the Orphan Drug Act, a panel established by the National Institutes of Health developed a set of Radioepidemiological Tables. The tables serve as a reference tool providing probability of causation estimates for individuals with cancer who were exposed to ionizing radiation. Use of the tables requires

[1]To regenerate our previous formula for Attributable Risk, define RadRisk to be $p_E - p_{NE}$.

information about the person's dose, gender, age at exposure, date of cancer diagnosis and other relevant factors. The tables are used by the Department of Veterans Affairs (DVA) to make compensation decisions for veterans with cancer who were exposed in the performance of duty to radiation from atomic weapon detonations.

The primary source of data for the 1985 tables is research on cancer-related deaths occurring among Japanese atomic bomb survivors from World War II.

The 1985 tables are presently being updated by the National Cancer Institute (NCI) and the Centers for Disease Control and Prevention to incorporate progress in research on the relationship between radiation and cancer risk. The draft update has been reviewed by the National Research Council and by NIOSH. DOL will employ the updated version of the tables, with modifications important to claims under EEOICPA, as a basis for determining probability of causation for employees covered under EEOICPA.

A major scientific change achieved by this update is the use of risk models developed from data on the occurrence of cancers (cases of illness) rather than the occurrence of cancer deaths among Japanese atomic bomb survivors. The risk models are further improved by being based on more current data as well. Many more cancers have been modeled in the revised report. The new risk models also take into account factors that modify the effect of radiation on cancer, related to the type of radiation dose, the amount of dose, and the timing of the dose.

A major technological change accompanying this update, which represents a scientific improvement, is the production of a computer software program for calculating probability of causation. This software program, named the Interactive RadioEpidemiological Program (IREP), allows the user to apply the NCI risk models directly to data on an individual employee. This makes it possible to estimate probability of causation using better quantitative methods than could be incorporated into printed tables. In particular, IREP allows the user to take into account uncertainty concerning the information being used to estimate probability of causation. There typically is uncertainty about the radiation dose levels to which a person has been exposed, as well as uncertainty relating levels of dose received to levels of cancer risk observed in study populations.

Accounting for uncertainty is important because it can have a large effect on the probability of causation estimates. DVA, in their use of the 1985 Radioepidemiological Tables, uses the probability of causation estimates found in the tables at the upper 99 percent credibility limit. This means when DVA determines whether the cancer of a veteran was more likely than not caused by radiation, they use the estimate that is 99 percent certain to be greater than the probability that would be calculated if the information on dose and the risk model were perfectly accurate. Similarly, these HHS guidelines, as required by EEOICPA, will use the upper 99 percent credibility limit to determine whether the cancers of employees are at least as likely as not caused by their occupational radiation doses. This will help minimize the possibility of denying compensation to claimants under EEOICPA for those employees with cancers likely to have been caused by occupational radiation exposures.

F. Use of IREP for Energy Employees

The risk models developed by NCI and CDC for IREP provide the primary basis for

developing guidelines for estimating probability of causation under EEOICPA. They directly address 33 cancers and most types of radiation exposure relevant to employees covered by EEOICPA. These models take into account the employee's cancer type, year of birth, year of cancer diagnosis, and exposure information such as years of exposure, as well as the dose received from gamma radiation, x-rays, alpha radiation, beta radiation, and neutrons during each year. Also, the risk model for lung cancer takes into account smoking history and the risk model for skin cancer takes into account race/ethnicity. None of the risk models explicitly accounts for exposure to other occupational, environmental, or dietary carcinogens. Models accounting for these factors have not been developed and may not be possible to develop based on existing research. Moreover, DOL could not consistently or efficiently obtain the data required to make use of such models.

IREP models do not specifically include cancers as defined in their early stages: *carcinoma in situ* (CIS). These lesions are becoming more frequently diagnosed, as the use of cancer screening tools, such as mammography, have increased in the general population. The risk factors and treatment for CIS are frequently similar to those for malignant neoplasms, and, while controversial, there is growing evidence that CIS represents the earliest detectable phase of malignancy. Therefore, for determining compensation under EEOICPA, HHS requires that CIS be treated as a malignant neoplasm of the specified site.

Cancers identified by their secondary sites (sites to which a malignant cancer has spread), when the primary site is unknown, raise another issue for the application of IREP. This situation will most commonly arise when death certificate information is the primary source of a cancer diagnosis. It is accepted in medicine that cancer causing agents such as ionizing radiation produce primary cancers. This means, in a case in which the primary site of cancer is unknown, the primary site must be established by inference to estimate probability of causation.

HHS establishes such assignments in these guidelines, based on an evaluation of the relationship between primary and secondary cancer sites using the National Center for Health Statistics (NCHS) Mortality Database for years 1995–1997. Because national cancer incidence databases (e.g., the National Cancer Institute's Surveillance, Epidemiology and End Results program) do not contain information about sites of metastasis, the NCHS database is the best available data source at this time to assign the primary site(s) most likely to have caused the spread of cancer to a known secondary site. For each secondary cancer, HHS identified the set of primary cancers producing approximately 75% of that secondary cancer among the U.S. population (males and females were considered separately). The sets are tabulated in this rule. DOL will determine the final assignment of a primary cancer site for an individual claim on a case-by-case basis, as the site among possible primary sites which results in the highest probability of causation estimate.

Employees diagnosed with two or more primary cancers also raise a special issue for determining probability of causation. Even under the assumption that the biological mechanisms by which each cancer is caused are unrelated, uncertainty estimates about the level of radiation delivered to each cancer site will be related. While fully understanding this situation requires statistical training, the consequence has simple but important implications. Under

this rule, instead of determining the probability that each cancer was caused by radiation independently, DOL will perform an additional statistical procedure following the use of IREP to determine the probability that at least one of the cancers was caused by the radiation. This approach is important to the claimant because it would determine a higher probability of causation than would be determined for either cancer individually. (pp. 22297–22298)

### 4.1.3 Appendix: District of Columbia Court of Appeals, "In Re As. H" (Decided: June 10, 2004)

DISTRICT OF COLUMBIA COURT OF APPEALS

In Re As. H.

SCHWELB, Associate Judge: This juvenile delinquency case is more than five years old. On January 20, 1999, following a factfinding hearing, As. H., then sixteen years of age, was adjudicated guilty of robbery. The sole evidence implicating As. H. in the offense was the testimony of the victim, Ms. Michal Freedhoff, who identified As. H. at a photo array almost a month after the robbery and again in court more than four months after that. Ms. Freedhoff described her level of certainty on both occasions, however, as "seven or eight" on a scale of one to ten. Because Ms. Freedhoff was obviously less than positive regarding her identification, and for other reasons described below, we conclude as a matter of law that the evidence was insufficient to prove beyond a reasonable doubt that As. H. was involved in the robbery. Accordingly, we reverse.

I. In the early morning hours of August 17, 1998, between 12:30 and 1:00 a.m., Ms. Freedhoff was robbed by three or more young men. The assailants knocked Ms. Freedhoff to the ground, threatened her with "a long piece of wood" which, Ms. Freedhoff believed, was "suppose[d] to look like a rifle," ordered her to "shut up, bitch," and robbed her of her purse and her personal electronic organizer. Ms. Freedhoff promptly reported the crime to the police. Officers detained a group of young men shortly after the robbery and arranged a show-up, but Ms. Freedhoff stated that the detained individuals were not the robbers. Indeed, she was "completely" certain that the individuals at the show-up were not the guilty parties.

Ms. Freedhoff testified that there were street lights in the area where the robbery occurred. She further stated that she had been outside in the street for some time, so that her eyes had become accustomed to the dark. Nevertheless, Ms. Freedhoff could not provide an informative description of her assailants. According to Detective Ross, she recalled nothing distinctive about their clothing; "young black males and baggy clothes" was his recollection of her report. At the factfinding hearing, which took place more than five months after the robbery, Ms. Freedhoff recalled that the robbers were teenagers, "two dark-skinned and one light," each of a different height, and that "one had shorts and sneakers and another may have had a hat." Ms. Freedhoff was also uncertain as to the role which the individual she tentatively identified as As. H. allegedly played in the robbery.

On September 11, 1998, Detective Ross showed Ms. Freedhoff an array of nine polaroid pictures and asked her if she recognized anyone who was involved in the offense. At a hearing

on As. H.'s motion to suppress identification, Ms. Freedhoff testified as follows regarding this array:

Q: Now, Ms. Freedhoff, on that day did you identify any of the people in the photos as having been involved in the incident of August 16th?

A: Yes, I did.

Q: Which photos did you identify?

A: These two marked nine and [ten] I was very certain about and the two marked three and four I was less certain about.

Q: During the identification procedure, did you talk to the detective about your level of certainty?

A: Yes.

Q: In terms of nine and [ten], what was your level of certainty that those people were involved?

A: I [was] asked to rate them on a scale of—I believe it was one to [ten]—and I believe I said it was, that nine and [ten], I was seven or eight.

Q: And in terms of three and four, how did you rate those?

A: Six.

According to Ms. Freedhoff, the photograph of As. H. was No. 10. At the factfinding hearing, Ms. Freedhoff initially stated that she saw one of the robbers sitting in the courtroom, pointing out As. H. When asked which of the individuals in the array he was, Ms. Freedhoff "believed" that it "would be Number 10." However, when counsel for the District of Columbia again asked Ms. Freedhoff about her present level of certainty in making the identification—how certain are you?—the witness adhered to her previous estimate: "At the time, on a scale of one to [ten], I said that I was seven or eight."

According to Detective Ross, who also testified regarding the viewing of the photo array, Ms. Freedhoff was "comfortable in saying they could be the people that robbed her." Ross further disclosed that he "may have discussed with [Ms. Freedhoff] that I had a previous history with the persons that she had picked. They were my possible suspects in the case."

Without elaborating on his reasons, the trial judge denied As. H.'s motion to suppress identification4 and found As. H. guilty as charged. This appeal followed.

II. In evaluating claims of evidentiary insufficiency in juvenile delinquency appeals, we view the record "in the light most favorable to the [District], giving full play to the right of the judge, as the trier of fact, to determine credibility, weigh the evidence, and draw reasonable inferences ... We will reverse on insufficiency grounds only when the [District] has failed to produce evidence upon which a reasonable mind might fairly find guilt beyond a reasonable doubt." "Even identification testimony of a single eyewitness will be sufficient so long as a reasonable person could find the identification convincing beyond a reasonable doubt." Moreover, the District was not required to prove As. H.'s guilt beyond all doubt. "There is no rule of law which requires an identification to be positive beyond any shadow of doubt."

Nevertheless, the "[beyond a] reasonable doubt" standard of proof is a formidable one. It "requires the factfinder to reach a subjective state of near certitude of the guilt of the

accused." Although appellate review is deferential, we have "the obligation to take seriously the requirement that the evidence in a criminal prosecution must be strong enough that a jury behaving rationally really could find it persuasive beyond a reasonable doubt." Moreover, "while [the trier of fact] is entitled to draw a vast range of reasonable inferences from evidence, [he or she] may not base [an adjudication of guilt] on mere speculation."[2]

In the present case, we have an eyewitness identification of questionable certitude, and the witness and the respondent are strangers. Ms. Freedhoff saw her assailants at night and under extremely stressful conditions. Moreover, this is a "pure" eyewitness identification case; there is no evidence linking As. H. to the robbery except for Ms. Freedhoff's statements upon viewing the array almost a month after the event and her testimony at the factfinding hearing more than five months after she was robbed.

The vagaries of eyewitness identification, and the potential for wrongful convictions or adjudications based upon such evidence, have long been recognized in the District of Columbia. More recently, in Webster v. United States, we summarized this concern as follows:

"[T]he identification of strangers is proverbially untrustworthy. The hazards of such testimony are established by a formidable number of instances in the records of English and American trials." FELIX FRANKFURTER, THE CASE OF SACCO AND VANZETTI (1927). Indeed, "[p]ositive identification of a person not previously known to the witness is perhaps the most fearful testimony known to the law of evidence." Even if the witness professes certainty, "it is well recognized that the most positive eyewitness is not necessarily the most reliable."

Here, the witness did not even profess certainty. Moreover, the present case concerns a hesitant cross-racial identification by a white woman of a black teenager, and "[i]t is well established that there exists a comparative difficulty in recognizing individual members of a race different from one's own." ELIZABETH LOFTUS, EYEWITNESS TESTIMONY; see State v. Cromedy, (discussing at length the difficulties in cross-racial identification and mandating a jury instruction on the subject in some cases); John P. Rutledge, They All Look Alike: The Inaccuracy of Cross-Racial Identifications, 28 AM. J. CRIM. L. 207 (2001).

It is in the context of these realities that we now turn to the dispositive issue in this appeal, namely, whether Ms. Freedhoff's testimony—the only evidence of As. H.'s participation in the robbery—was legally sufficient to support a finding of guilt beyond a reasonable doubt. The key fact is that, both when viewing the polaroid photographs and when testifying in open court, Ms. Freedhoff candidly characterized her level of "certainty"—i.e., of her being "very certain"—as seven or eight on a scale of one to ten. Her testimony leads inexorably to the conclusion that her level of uncertainty—i.e., the possibility that As. H. was not involved—was two or three out of ten—a 20% to 30% possibility of innocence. This differs dramatically

[2]The court emphasized in Crawley that as appellate judges, we have the responsibility in eyewitness identification cases "to draw upon our own experience, value judgments, and common sense in determining whether the [finding] reached was in keeping with the facts." Although this observation might be viewed today as an unduly activist formulation of an appellate court's function, it illustrates the concern of conscientious judges regarding the possibility that a mistaken identification may send an innocent person to prison.

from Ms. Freedhoff's complete certainty that the young men she viewed at the show-up on the night of the offense were not the robbers. The contrast between Ms. Freedhoff's statements in the two situations is revealing, and surely negates the "near certitude" that is required for a showing of guilt beyond a reasonable doubt. The "seven or eight out of ten" assessment is also consistent with Detective Ross' recollection of Ms. Freedhoff's account: As. H. and others "could be the people that robbed her," and As. H. "looked like" one of the kids. It is, of course, difficult (if not impossible) to place a meaningful numerical value on reasonable doubt. See generally Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, 84 HARV. L. REV 1329 (1971); Underwood, The Thumb on the Scales of Justice; Burden of Persuasion in Criminal Cases, 86 Yale L.J. 1299, 1309–11 (1977) (hereinafter Underwood). Professor Wigmore cites a study in which judges in Chicago were asked to:
translate into probability statements their sense of what it means to be convinced by a preponderance of the evidence, and to be convinced beyond a reasonable doubt. When responding to questionnaires, at least, the judges thought there was an important difference: almost a third of the responding judges put "beyond a reasonable doubt" at 100%, another third put it at 90% or 95%, and most of the rest put it at 80% or 85%. For the preponderance standard, by contrast, over half put it at 55%, and most of the rest put it between 60% and 75%.

Although the Chicago study alone is not dispositive of this appeal, it reveals that very few judges, if any, would have regarded an 80% probability as sufficient to prove guilt beyond a reasonable doubt, and that all of them would have considered a 70% probability as altogether inadequate. For the Chicago judges, Ms. Freedhoff's "certainty" appears to be well outside the ballpark for proof in a criminal case. In Fatico, nine judges of the United States District Court for the Eastern District of New York, responding to a poll by Judge Weinstein, the co-author of a leading text on evidence, suggested percentages of 76%, 80%, 85%, 85%, 85%, 85%, 90%, 90% and 95%, as reflecting the standard for proof beyond a reasonable doubt. Thus, at most, two of the nine judges polled by Judge Weinstein would have found the level of assurance voiced by Ms. Freedhoff sufficient to support a finding of guilt.[3]

But, argues the District, Ms. Freedhoff "was not asked for a level of accuracy or how sure she was, but, given the certainty of her identification, how high a level of certainty she had felt." Therefore, the argument goes, "the trier of fact can be confident that the witness felt that her identification was very certain." We do not find this contention at all persuasive. Taken to its logical conclusion, it would mean that if Ms. Freedhoff had expressed a level of certainty of one in ten—10%—this would be sufficient to support a finding of guilt. The notion that Ms. Freedhoff was assessing varying gradations of certainty, all of them very certain, is also at odds with what she told Detective Ross, namely, that As. H. "looks like"

[3]Commenting on the same Chicago study in one of its submissions, the District reveals only that "about one-third of the judges put it at 80%-85%." Unfortunately, by failing to mention that one third of the judges put "beyond a reasonable doubt" at 100% and that another third put it at 90%-95%, the District presents us with a misleading picture of the results of the study. Remarkably, the District then argues that we should affirm because judges who try to quantify reasonable doubt place it "not that far distant from Ms. Freedhoff's estimate." In fact, the contrast between the judges' estimates and Ms. Freedhoff's articulation is quite remarkable, and a contention that fails to take this contrast into account is necessarily fallacious.

or "could have been" one of the robbers.[4]

Professor Lawrence Tribe has written:

[I]t may well be ... that there is something intrinsically immoral about condemning a man as a criminal while telling oneself, "I believe that there is a chance of one in twenty that this defendant is innocent, but a 1/20 risk of sacrificing him erroneously is one I am willing to run in the interest of the public's—and my own—safety."

It may be that Professor Tribe's proposition is more suited to the world of academe than to the less rarefied realities of the Superior Court's criminal docket—realities in which "beyond all doubt" presents an idealist's impossible dream, while "beyond a reasonable doubt" provides a workable standard. This case, however, is not like the hypothetical one that disturbed Professor Tribe. Here, the doubt of the sole identifying witness in a night-time robbery by strangers to her stood at two or three out of ten, or 20%-30%. We conclude, at least on this record, that this level of uncertainty constituted reasonable doubt as a matter of law. Accordingly, we reverse the adjudication of guilt and remand the case to the Superior Court with directions to enter a judgment of not guilty and to dismiss the petition.[5]

So ordered.

FARRELL, Associate Judge, dissenting: Less than a month after she was assaulted by three young men, the complainant, Ms. Freedhoff, identified two men from photographs as among the assailants. One was appellant. According to the detective who showed her the photographs, she did not hesitate in picking appellant, and at the hearing on appellant's motion to suppress the identification she twice stated that she had been "very certain" in selecting his photograph. At trial, although she could not remember appellant's exact role in the assault, she stated that she had been able to see all three assailants well, that the two people she was "certain of" in her identification "were probably the two" who had been "in front of [her]" during the assault, and that she had identified them because "they looked very familiar to [her] as being the people that were involved." Ms. Freedhoff was not given

[4]The District also argues that, in open court, Ms. Freedhoff "unhesitatingly and positively" identified the respondent. As we have explained in Part I of this opinion, however, the full context of Ms. Freedhoff's courtroom testimony reveals that, five months after the robbery, she was no more certain of her identification than she had been when she viewed the photo array. Moreover, after Ms. Freedhoff had selected photographs at the array, Detective Ross revealed that he "had a previous history with the persons she had picked," and that they were his "possible suspects in the case." "[W]here ... the police consider an individual to be a possible perpetrator and a witness makes an initially ambiguous identification, there may develop a process of mutual bolstering which converts initial tentativeness into ultimate certainty." "The victim relies on the expertise of the officer and the officer upon the victim's identification."

[5]Our dissenting colleague argues that reasonable doubt is not susceptible of ready quantification, and we agree. But where, as in this case, the sole identifying witness described her own level of "certainty" as only seven or eight on a scale of ten, then, notwithstanding the difficulty of quantification in the abstract, this level of unsureness necessarily raises a reasonable doubt and negates the requisite finding of "near certitude" that As. H. was one of the robbers. Nothing in this opinion holds or even remotely suggests that a cross-racial identification is insufficient as a matter of law or that the trier of fact is required to discount such an identification. The reasonable doubt in this case arises from the witness' very limited certainty (seven or eight on a scale of ten) regarding her uncorroborated identification. The difficulties of eyewitness identification of strangers in general, as well as of cross-racial identification, provide the context in which the witness' uncertainty arose.

to quick accusations: at a show-up confrontation shortly after the assault, she had been "[completely] certain" that the individuals shown to her were not the assailants. The trial judge, sitting as trier of fact, found her testimony convincing and found appellant guilty beyond a reasonable doubt.

The majority sets that verdict aside. Although concededly unable to replicate Judge Mitchell's vantage point in assessing the complainant's demeanor and the strength of her belief as she recalled the robbery and identification, it concludes that the identification was too weak as a matter of law to support conviction. And it does so at bottom for one reason: when asked by the detective her level of certainty "on a scale of one to ten" in identifying appellant, Ms. Freedhoff had answered "seven or eight." This, in the majority's view, explains what she meant when she said she was "very certain," and a level of uncertainty of an uncorroborated eyewitness "st[anding] at two or three out of ten, or 20%-30%[,] ... constituted reasonable doubt as a matter of law."

The majority thus decides that the trier of fact could not convict based on testimony of a victim who was as much as four-fifths certain of her identification. I do not agree, basically because I believe that the entire effort to quantify the standard of proof beyond a reasonable doubt is a search for fool's gold. Ms. Freedhoff stated that she was very certain of her identification; she was questioned extensively about the circumstances of the photo display and the assault; and she offered reasons for her certainty. The fact that when asked to rate her certainty "on a scale of one to ten" she answered "seven or eight" cannot be decisive unless, like the majority, one is ready to substitute an unreliable, quantitative test of certainty for the intensely qualitative standard of proof beyond a reasonable doubt. Even in popular usage, the "scale of one to ten" as an indicator of belief is notoriously imprecise. People who in any ultimate—and unascertainable—sense probably share the same level of conviction may translate that very differently into numbers, and even the same person will change his mind from one moment to the next in assigning a percentage to his belief. Treating "one to ten" as a decisive indicator of the sufficiency of identification evidence thus elevates to a legal standard a popular measure that makes no claim at all to precision. As Wigmore stated long ago in this context, "The truth is that no one has yet invented or discovered a mode of measurement for the intensity of human belief. Hence there can be yet no successful method of communicating intelligibly ... a sound method of self-analysis for one's belief." Here, for example, Ms. Freedhoff equated "seven or eight" with being "very certain"; for all we know, she thought that any higher number would approach mathematical or absolute certainty, something the reasonable doubt standard does not require. The trial judge wisely did not view her attempt to furnish a numerical equivalent for her belief as conclusive, and neither should we.

The judicial straw polls cited by the majority merely confirm the futility of defining a percentual range (or "ball-park," to quote the majority) within which proof beyond a reasonable doubt must lie. Had Ms. Freedhoff added five percent to her belief-assessment (as much as "85%" rather than as much as "80%"), she would have come well within the range of, for example, Judge Weinstein's survey in Fatico. A factfinder's evaluation of credibility and intensity of belief should not be overridden by such inexact and even trivial differences

of quantification.

Another aspect of the majority's opinion requires comment. It points to "[t]he vagaries of eyewitness identification," explains that this was a case of "cross-racial identification by a white woman of a black teenager," and cites to the "well established ... comparative difficulty in recognizing individual members of a race different from one's own." [quoting ELIZABETH LOFTUS, EYEWITNESS TESTIMONY]. It is not clear what the majority means by this discussion. The present appeal is not about whether a trier of fact may hear expert testimony or be instructed regarding the uncertainties of eyewitness identification, cross-racial or any other. Here the majority holds the identification insufficient as a matter of law, which implies that the trier of fact was required to discount the identification to an (undefined) extent because of the intrinsic weakness of eyewitness identifications generally or because this one was cross-racial. Either basis would be unprecedented. If, as I prefer to believe, that is not what the majority intends, then I respectfully suggest that the entire discussion of the point is dictum.

I would affirm the judgment of the trial court

## 4.2 Suggested Reading on Issues of Risk

In making medical decisions, either for ourselves or as suggested practice more generally, there are always tradeoffs between benefits and risks. Although not usually definable in a random variable context where a choice might be made based on whether the game was fair or unfair in our favor, some type of weighing of possible outcomes is still necessary. In taking a drug, for example, we might consider the outcome of "getter better" against that of having an "adverse event"—think Vioxx and a heart attack. As we are all painfully aware, there is continual (24/7) media coverage for these kinds of considerations. A fairly large collection is listed in the Suggested Reading (Section B.6) on various risk/benefit tradeoffs that we might have faced or may need to face in the future.

B.6 – Risk Issues

B.6.1 – When Lowering the Odds of Cancer Isn't Enough (Tara Parker-Pope, *New York Times*, December 14, 2009)

B.6.2 – After Cancer, Removing a Healthy Breast (Tara Parker-Pope and Stuart Bradford, *New York Times*, March 8, 2010)

B.6.3 – Assessing Your Risk of a Heart Attack (*New York Times*, March 10, 2010)

B.6.4 – Panel Urges New Look at Caesarean Guidelines (Denise Grady, *New York Times*, March 10, 2010)

B.6.5 – Caesarean Births Are at a High in U.S. (Denise Grady, *New York Times*, March 23, 2010)

B.6.6 – Risks Seen in Cholesterol Drug Use in Healthy People (Duff Wilson, *New York Times*, March 30, 2010)

B.6.7 – In Medicine, the Power of No (David Leonhardt, *New York Times*, April 6, 2010)

B.6.8 – Perceptions of Risk Can Get Pulled in Opposite Directions (Bruce Bower, *Science News*, April 24, 2010)

B.6.9 – Hospital Infection Problem Persists (Kevin Sack, *New York Times*, April 13, 2010)

B.6.10 – Drivers and Legislators Dismiss Cellphone Risks (Matt Richtel, *New York Times*, July 19, 2009)

B.6.11 – It Started More Than One Revolution (Gardiner Harris, *New York Times*, May 3, 2010)

B.6.12 – Diabetes Heart Treatments May Cause Harm (Gina Kolata, *New York Times*, March 14, 2010)

B.6.13 – F.D.A. Steps Up Oversight of Infusion Pumps (Barry Meier, *New York Times*, April 23, 2010)

B.6.14 – Lung Disease: A Human Cost of 'Worn' Denim (Janet Raloff, *Science News*, April 18, 2010)

B.6.15 – Vitamin D: Obese and 'Uniform' Risks (Janet Raloff, *Science News*, April 29, 2010)

B.6.16 – IOM: Manufacturers Should Help Americans Cut Back on Salt (Janet Raloff, *Science News*, April 20, 2010)

B.6.17 – U.S. Panel Criticized as Overstating Cancer Risks (Denise Grady, *New York Times*, May 6, 2010)

B.6.18 – Study Pokes Holes in Air Bag Standards (Jo Craven McGinty, *New York Times*, May 14, 2010)

B.6.19 – Cass Sunstein Wants to Nudge Us (Benjamin Wallace-Wells, *New York Times*, May 11, 2010)

B.6.20 – Whose Body is it, Anyway? (Atul Gawande, *New Yorker*, October 4, 1999 – excerpted by Meenal and Bashir Mamdani for the *Indian Journal of Medical Ethics*)

## 4.3 Suggested Reading on Issues of Betting and Gaming

In addition to issues of personal medical risk, we regularly face a variety of betting and gambling situations. Some of these involve buying insurance, financial investments, and the value of information generally in making informed decisions. We list additional Suggested Reading on these topics in Section B.7 (Betting and Gaming)

B.7 – Betting and Gaming

B.7.1 – Stock Traders Find Speed Pays, in Milliseconds (Charles Duhigg, *New York Times*, July 24, 2009)

B.7.2 – Bust the Health Care Trusts (Robert B. Reich, *New York Times*, February 24, 2010)

B.7.3 – How Men's Overconfidence Hurts Them as Investors (Jeff Sommer, *New York Times*, March 12, 2010)

B.7.4 – A Routine Deal Became an $840 Million Mistake (Landon Thoman Jr., *New York Times*, April 22, 2010)

B.7.5 – Fortunate Timing Seals a Deal (Floyd Norris, *New York Times*, April 22, 2010)

B.7.6 – Don't Cry for Wall Street (Paul Krugman, *New York Times*, April 23, 2010)

B.7.7 – Goldman Cited 'Serious' Profit on Mortgages (Louise Story and Sewell Chan, *New York Times*, April 24, 2010)

B.7.8 – Rating Agency Data Aided Wall Street in Deals (Gretchen Morgenson and Louise Story, *New York Times*, April 23, 2010)

B.7.9 – Fight On, Goldman Sachs! (Frank Rich, *New York Times*, April 25, 2010)

B.7.10 – Prosecutors Ask if 8 Banks Duped Rating Agencies (Louise Story, *New York Times*, May 12, 2010)

B.7.11 – Clients Worried About Goldman's Dueling Goals (Gretchen Morgenson and Louise Story, *New York Times*, May 18, 2010)

B.7.12 – Naked Truth on Default Swaps (Floyd Norris, *New York Times*, May 20, 2010)

B.7.13 – Suddenly, the Rating Agencies Don't Look Untouchable (David Segal, *New York Times*, May 21, 2010)

B.7.14 – Fixing Wall Street's Autopilot (Michael Durbin, *New York Times*, May 7, 2010)

# Chapter 5

# Correlation

# Chapter 6

# Prediction

### 6.0.1 Appendix: Continuation of the American Psychiatric Association, *Amicus Curiae* Brief: Barefoot v. Estelle

INTRODUCTION AND SUMMARY OF ARGUMENT

The questions presented in this case are the logical outgrowth of two prior decisions by this Court. In the first, Jurek v. Texas, the Court dealt with the same Texas capital sentencing procedure involved here. The Court there rejected a constitutional challenge to the "future dangerousness" question, ruling that the statutory standard was not impermissibly vague. Although recognizing the difficulty inherent in predicting future behavior, the Court held that "[t]he task that [the] jury must perform ... is basically no different from the task performed countless times each day throughout the American system of criminal justice." The Jurek Court thus upheld the use of the Texas statutory question, but did not consider the types of evidence that could be presented to the jury for purposes of this determination.

Subsequently in Estelle v. Smith, the Court again dealt with the Texas sentencing scheme—this time in the context of a psychiatric examination to determine the defendant's competency to stand trial. The Court held that the Fifth Amendment's privilege against self-incrimination applied to such psychiatric examinations, at least to the extent that a prosecution psychiatrist later testifies concerning the defendant's future dangerousness. The Court reasoned that although a defendant has no generalized constitutional right to remain silent at a psychiatric examination properly limited to the issues of sanity or competency, full Miranda warnings must be given with respect to testimony concerning future dangerousness because of "the gravity of the decision to be made at the penalty phase ... " The Smith decision thus enables a capital defendant to bar a government psychiatric examination on the issue of future dangerousness.

The [present] case raises the two issues left unresolved in Jurek and Smith. These are, first, whether a psychiatrist, testifying as an expert medical witness, may ever be permitted to render a prediction as to a capital defendant's long-term future dangerousness. The second issue is whether such testimony may be elicited on the basis of hypothetical questions, even if there exists no general prohibition against the use of expert psychiatric testimony on the

issue of long-term future dangerousness. *Amicus* believes that both of these questions should be answered in the negative.

I. Psychiatrists should not be permitted to offer a prediction concerning the long-term future dangerousness of a defendant in a capital case, at least in those circumstances where the psychiatrist purports to be testifying as a medical expert possessing predictive expertise in this area. Although psychiatric assessments may permit short-term predictions of violent or assaultive behavior, medical knowledge has simply not advanced to the point where long-term predictions—the type of testimony at issue in this case—may be made with even reasonable accuracy. The large body of research in this area indicates that, even under the best of conditions, psychiatric predictions of long-term future dangerousness are wrong in at least two out of every three cases.

The forecast of future violent conduct on the part of a defendant in a capital case is, at bottom, a lay determination, not an expert psychiatric determination. To the extent such predictions have any validity, they can only be made on the basis of essentially actuarial data to which psychiatrists, qua psychiatrists, can bring no special interpretative skills. On the other hand, the use of psychiatric testimony on this issue causes serious prejudice to the defendant. By dressing up the actuarial data with an "expert" opinion, the psychiatrist's testimony is likely to receive undue weight. In addition, it permits the jury to avoid the difficult actuarial questions by seeking refuge in a medical diagnosis that provides a false aura of certainty. For these reasons, psychiatric testimony on future dangerousness impermissibly distorts the fact-finding process in capital cases.

II. Even if psychiatrists under some circumstances are allowed to render an expert medical opinion on the question of future dangerousness, *amicus* submits that they should never be permitted to do so unless they have conducted a psychiatric examination of the defendant. It is evident from the testimony in this case that the key clinical determination relied upon by both psychiatrists was their diagnosis of "sociopathy" or "antisocial personality disorder." However, such a diagnosis simply cannot be made on the basis of a hypothetical question. Absent an in-depth psychiatric examination and evaluation, the psychiatrist cannot exclude alternative diagnoses; nor can he assure that the necessary criteria for making the diagnosis in question are met. As a result, he is unable to render a medical opinion with a reasonable degree of certainty.

These deficiencies strip the psychiatric testimony of all value in the present context. Even assuming that the diagnosis of antisocial personality disorder is probative of future dangerousness—an assumption which we do not accept—it is nonetheless clear that the limited facts given in the hypothetical fail to disprove other illnesses that plainly do not indicate a general propensity to commit criminal acts. Moreover, these other illnesses may be more amenable to treatment—a factor that may further reduce the likelihood of future aggressive behavior by the defendant.

. . .

### 6.0.2 Appendix: Opinion and Dissent in the U.S. Supreme Court, Barefoot v. Estelle (Decided, July 6, 1983)

Summary of the majority opinion:

(a) There is no merit to petitioner's argument that psychiatrists, individually and as a group, are incompetent to predict with an acceptable degree of reliability that a particular criminal will commit other crimes in the future, and so represent a danger to the community. To accept such an argument would call into question predictions of future behavior that are constantly made in other contexts. Moreover, under the generally applicable rules of evidence covering the admission and weight of unprivileged evidence, psychiatric testimony predicting dangerousness may be countered not only as erroneous in a particular case but also as generally so unreliable that it should be ignored. Nor, despite the view of the American Psychiatric Association supporting petitioner's view, is there any convincing evidence that such testimony is almost entirely unreliable, and that the factfinder and the adversary system will not be competent to uncover, recognize, and take due account of its shortcomings.

(b) Psychiatric testimony need not be based on personal examination of the defendant, but may properly be given in response to hypothetical questions. Expert testimony, whether in the form of an opinion based on hypothetical questions or otherwise, is commonly admitted as evidence where it might help the factfinder do its job. Although this case involves the death penalty, there is no constitutional barrier to applying the ordinary rules of evidence governing the use of expert testimony.

. . .

Justice Blackmun dissenting:

I agree with most of what Justice Marshall has said in his dissenting opinion. I, too, dissent, but I base my conclusion also on evidentiary factors that the Court rejects with some emphasis. The Court holds that psychiatric testimony about a defendant's future dangerousness is admissible, despite the fact that such testimony is wrong two times out of three. The Court reaches this result—even in a capital case—because, it is said, the testimony is subject to cross-examination and impeachment. In the present state of psychiatric knowledge, this is too much for me. One may accept this in a routine lawsuit for money damages, but when a person's life is at stake—no matter how heinous his offense—a requirement of greater reliability should prevail. In a capital case, the specious testimony of a psychiatrist, colored in the eyes of an impressionable jury by the inevitable untouchability of a medical specialist's words, equates with death itself.

To obtain a death sentence in Texas, the State is required to prove beyond a reasonable doubt that "there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society." As a practical matter, this prediction of future dangerousness was the only issue to be decided by Barefoot's sentencing jury.

At the sentencing hearing, the State established that Barefoot had two prior convictions for drug offenses and two prior convictions for unlawful possession of firearms. None of these convictions involved acts of violence. At the guilt stage of the trial, for the limited purpose of establishing that the crime was committed in order to evade police custody, the State had

presented evidence that Barefoot had escaped from jail in New Mexico where he was being held on charges of statutory rape and unlawful restraint of a minor child with intent to commit sexual penetration against the child's will. The prosecution also called several character witnesses at the sentencing hearing, from towns in five States. Without mentioning particular examples of Barefoot's conduct, these witnesses testified that Barefoot's reputation for being a peaceable and law-abiding citizen was bad in their respective communities.

Last, the prosecution called Doctors Holbrook and Grigson, whose testimony extended over more than half the hearing. Neither had examined Barefoot or requested the opportunity to examine him. In the presence of the jury, and over defense counsel's objection, each was qualified as an expert psychiatrist witness. Doctor Holbrook detailed at length his training and experience as a psychiatrist, which included a position as chief of psychiatric services at the Department of Corrections. He explained that he had previously performed many "criminal evaluations," and that he subsequently took the post at the Department of Corrections to observe the subjects of these evaluations so that he could "be certain those opinions that [he] had were accurate at the time of trial and pretrial." He then informed the jury that it was "within [his] capacity as a doctor of psychiatry to predict the future dangerousness of an individual within a reasonable medical certainty,", and that he could give

"an expert medical opinion that would be within reasonable psychiatric certainty as to whether or not that individual would be dangerous to the degree that there would be a probability that that person would commit criminal acts of violence in the future that would constitute a continuing threat to society."

Doctor Grigson also detailed his training and medical experience, which, he said, included examination of "between thirty and forty thousand individuals," including 8,000 charged with felonies, and at least 300 charged with murder. He testified that, with enough information, he would be able to "give a medical opinion within reasonable psychiatric certainty as to the psychological or psychiatric makeup of an individual," and that this skill was "particular to the field of psychiatry, and not to the average layman."

Each psychiatrist then was given an extended hypothetical question asking him to assume as true about Barefoot the four prior convictions for nonviolent offenses, the bad reputation for being law-abiding in various communities, the New Mexico escape, the events surrounding the murder for which he was on trial and, in Doctor Grigson's case, the New Mexico arrest. On the basis of the hypothetical question, Doctor Holbrook diagnosed Barefoot "within a reasonable psychiatr[ic] certainty," as a "criminal sociopath." He testified that he knew of no treatment that could change this condition, and that the condition would not change for the better but "may become accelerated" in the next few years. Finally, Doctor Holbrook testified that, "within reasonable psychiatric certainty," there was "a probability that the Thomas A. Barefoot in that hypothetical will commit criminal acts of violence in the future that would constitute a continuing threat to society," and that his opinion would not change if the "society" at issue was that within Texas prisons, rather than society outside prison.

Doctor Grigson then testified that, on the basis of the hypothetical question, he could diagnose Barefoot "within reasonable psychiatric certainty" as an individual with "a fairly

classical, typical, sociopathic personality disorder." He placed Barefoot in the "most severe category of sociopaths (on a scale of one to ten, Barefoot was "above ten"), and stated that there was no known cure for the condition. Finally, Doctor Grigson testified that whether Barefoot was in society at large or in a prison society there was a "one hundred percent and absolute" chance that Barefoot would commit future acts of criminal violence that would constitute a continuing threat to society.

On cross-examination, defense counsel questioned the psychiatrists about studies demonstrating that psychiatrists' predictions of future dangerousness are inherently unreliable. Doctor Holbrook indicated his familiarity with many of these studies, but stated that he disagreed with their conclusions. Doctor Grigson stated that he was not familiar with most of these studies, and that their conclusions were accepted by only a "small minority group" of psychiatrists—"[i]t's not the American Psychiatric Association that believes that.

After an hour of deliberation, the jury answered "yes" to the two statutory questions, and Thomas Barefoot was sentenced to death.

The American Psychiatric Association (APA), participating in this case as *amicus curiae*, informs us that "[t]he unreliability of psychiatric predictions of long-term future dangerousness is by now an established fact within the profession." The APA's best estimate is that two out of three predictions of long-term future violence made by psychiatrists are wrong. The Court does not dispute this proposition, and indeed it could not do so; the evidence is overwhelming. For example, the APA's Draft Report of the Task Force on the Role of Psychiatry in the Sentencing Process (1983) states that

"[c]onsiderable evidence has been accumulated by now to demonstrate that long-term prediction by psychiatrists of future violence is an extremely inaccurate process."

John Monahan, recognized as "the leading thinker on this issue" even by the State's expert witness at Barefoot's federal habeas corpus hearing, concludes that

"the 'best' clinical research currently in existence indicates that psychiatrists and psychologists are accurate in no more than one out of three predictions of violent behavior,"

even among populations of individuals who are mentally ill and have committed violence in the past. Another study has found it impossible to identify any subclass of offenders "whose members have a greater-than-even chance of engaging again in an assaultive act." Yet another commentator observes:

"In general, mental health professionals ... are more likely to be wrong than right when they predict legally relevant behavior. When predicting violence, dangerousness, and suicide, they are far more likely to be wrong than right."

Neither the Court nor the State of Texas has cited a single reputable scientific source contradicting the unanimous conclusion of professionals in this field that psychiatric predictions of long-term future violence are wrong more often than they are right.

The APA also concludes, as do researchers that have studied the issue, that psychiatrists simply have no expertise in predicting long-term future dangerousness. A layman with access to relevant statistics can do at least as well, and possibly better; psychiatric training is not relevant to the factors that validly can be employed to make such predictions, and psychiatrists consistently err on the side of overpredicting violence. Thus, while Doctors

Grigson and Holbrook were presented by the State and by self-proclamation as experts at predicting future dangerousness, the scientific literature makes crystal clear that they had no expertise whatever. Despite their claims that they were able to predict Barefoot's future behavior "within reasonable psychiatric certainty," or to a "one hundred percent and absolute" certainty, there was, in fact, no more than a one in three chance that they were correct.[1]

It is impossible to square admission of this purportedly scientific but actually baseless testimony with the Constitution's paramount concern for reliability in capital sentencing.[2] Death is a permissible punishment in Texas only if the jury finds beyond a reasonable doubt that there is a probability the defendant will commit future acts of criminal violence. The admission of unreliable psychiatric predictions of future violence, offered with unabashed

[1]Like the District Court ... and the Court of Appeals, ... the Court seeks to justify the admission of psychiatric testimony on the ground that

"[t]he majority of psychiatric experts agree that where there is a pattern of repetitive assaultive and violent conduct, the accuracy of psychiatric predictions of future dangerousness dramatically rises."

... The District Court correctly found that there is empirical evidence supporting the common sense correlation between repetitive past violence and future violence; the APA states that

"[t]he most that can be said about any individual is that a history of past violence increases the probability that future violence will occur."

But psychiatrists have no special insights to add to this actuarial fact, and a single violent crime cannot provide a basis for a reliable prediction of future violence. ...

The lower courts and this Court have sought solace in this statistical correlation without acknowledging its obvious irrelevance to the facts of this case. The District Court did not find that the State demonstrated any pattern of repetitive assault and violent conduct by Barefoot. Recognizing the importance of giving some credibility to its experts' specious prognostications, the State now claims that the "reputation" testimony adduced at the sentencing hearing "can only evince repeated, widespread acts of criminal violence." ... This is simply absurd. There was no testimony worthy of credence that Barefoot had committed acts of violence apart from the crime for which he was being tried; there was testimony only of a bad reputation for peaceable and law-abiding conduct. In light of the fact that each of Barefoot's prior convictions was for a nonviolent offense, such testimony obviously could have been based on antisocial but nonviolent behavior. Neither psychiatrist informed the jury that he considered this reputation testimony to show a history of repeated acts of violence. Moreover, if the psychiatrists or the jury were to rely on such vague hearsay testimony in order to show a "pattern of repetitive assault and violent conduct," Barefoot's death sentence would rest on information that might "bear no closer relation to fact than the average rumor or item of gossip," ... and should be invalid for that reason alone. A death sentence cannot rest on highly dubious predictions secretly based on a factual foundation of hearsay and pure conjecture. ...

[2]Although I believe that the misleading nature of any psychiatric prediction of future violence violates due process when introduced in a capital sentencing hearing, admitting the predictions in this case—which were made without even examining the defendant—was particularly indefensible. In the APA's words, if prediction following even an in-depth examination is inherently unreliable,

"there is all the more reason to shun the practice of testifying without having examined the defendant at all. ... Needless to say, responding to hypotheticals is just as fraught with the possibility of error as testifying in any other way about an individual whom one has not personally examined. Although the courts have not yet rejected the practice, psychiatrists should."

... Such testimony is offensive not only to legal standards; the APA has declared that "[i]t is unethical for a psychiatrist to offer a professional opinion unless he/she has conducted an examination." ... The Court today sanctions admission in a capital sentencing hearing of "expert" medical testimony so unreliable and unprofessional that it violates the canons of medical ethics.

claims of "reasonable medical certainty" or "absolute" professional reliability, creates an intolerable danger that death sentences will be imposed erroneously.

The plurality in Woodson v. North Carolina, stated:

"Death, in its finality, differs more from life imprisonment than a 100-year prison term differs from one of only a year or two. Because of that qualitative difference, there is a corresponding difference in the need for reliability in the determination that death is the appropriate punishment in a specific case."

The Court does not see fit to mention this principle today, yet it is as firmly established as any in our Eighth Amendment jurisprudence. Only two weeks ago, in Zant v. Stephens, the Court described the need for reliability in the application of the death penalty as one of the basic "themes . . . reiterated in our opinions discussing the procedures required by the Constitution in capital sentencing determinations." (capital punishment must be "imposed fairly, and with reasonable consistency, or not at all"). State evidence rules notwithstanding, it is well established that, because the truth-seeking process may be unfairly skewed, due process may be violated even in a noncapital criminal case by the exclusion of evidence probative of innocence, or by the admission of certain categories of unreliable and prejudicial evidence ("[i]t is the reliability of identification evidence that primarily determines its admissibility"). The reliability and admissibility of evidence considered by a capital sentencing factfinder is obviously of still greater constitutional concern.

The danger of an unreliable death sentence created by this testimony cannot be brushed aside on the ground that the "jury [must] have before it all possible relevant information about the individual defendant whose fate it must determine." Although committed to allowing a "wide scope of evidence" at presentence hearings, the Court has recognized that "consideration must be given to the quality, as well as the quantity, of the information on which the sentencing [authority] may rely." Thus, very recently, this Court reaffirmed a crucial limitation on the permissible scope of evidence: "[s]o long as the evidence introduced . . . do[es] not prejudice a defendant, it is preferable not to impose restrictions." The Court all but admits the obviously prejudicial impact of the testimony of Doctors Grigson and Holbrook; granting that their absolute claims were more likely to be wrong than right, the Court states that "[t]here is no doubt that the psychiatric testimony increased the likelihood that petitioner would be sentenced to death.". Indeed, unreliable scientific evidence is widely acknowledged to be prejudicial. The reasons for this are manifest. "The major danger of scientific evidence is its potential to mislead the jury; an aura of scientific infallibility may shroud the evidence, and thus lead the jury to accept it without critical scrutiny."[3]

[3]There can be no dispute about this obvious proposition:

"Scientific evidence impresses lay jurors. They tend to assume it is more accurate and objective than lay testimony. A juror who thinks of scientific evidence visualizes instruments capable of amazingly precise measurement, of findings arrived at by dispassionate scientific tests. In short, in the mind of the typical lay juror, a scientific witness has a special aura of credibility."

. . . "Scientific . . . evidence has great potential for misleading the jury. The low probative worth can often be concealed in the jargon of some expert . . . ". This danger created by use of scientific evidence frequently has been recognized by the courts. Speaking specifically of psychiatric predictions of future dangerousness similar to those at issue, one District Court has observed that, when such a prediction

Where the public holds an exaggerated opinion of the accuracy of scientific testimony, the prejudice is likely to be indelible. There is little question that psychiatrists are perceived by the public as having a special expertise to predict dangerousness, a perception based on psychiatrists' study of mental disease. It is this perception that the State in Barefoot's case sought to exploit. Yet mental disease is not correlated with violence, and the stark fact is that no such expertise exists. Moreover, psychiatrists, it is said, sometimes attempt to perpetuate this illusion of expertise, and Doctors Grigson and Holbrook—who purported to be able to predict future dangerousness "within reasonable psychiatric certainty," or absolutely—present extremely disturbing examples of this tendency. The problem is not uncommon.

Furthermore, as is only reasonable, the Court's concern in encouraging the introduction of a wide scope of evidence has been to ensure that accurate information is provided to the sentencing authority without restriction. The joint opinion announcing the judgment in Gregg explained the jury's need for relevant evidence in these terms:

"If an experienced trial judge, who daily faces the difficult task of imposing sentences, has a vital need for accurate information ... to be able to impose a rational sentence in the typical criminal case, then accurate sentencing information is an indispensable prerequisite to a reasoned determination of whether a defendant shall live or die by a jury of people who may never before have made a sentencing decision."

So far as I am aware, the Court never has suggested that there is any interest in providing deceptive and inaccurate testimony to the jury. Psychiatric predictions of future dangerousness are not accurate; wrong two times out of three, their probative value, and therefore any possible contribution they might make to the ascertainment of truth, is virtually nonexistent (psychiatric testimony not sufficiently reliable to support finding that individual will be dangerous under any standard of proof). Indeed, given a psychiatrist's prediction that an individual will be dangerous, it is more likely than not that the defendant will not commit further violence. It is difficult to understand how the admission of such predictions can be justified as advancing the search for truth, particularly in light of their clearly prejudicial effect. Thus, the Court's remarkable observation that "[n]either petitioner nor the [APA] suggests that psychiatrists are always wrong with respect to future dangerousness, only most of the time," misses the point completely, and its claim that this testimony was no more problematic than "other relevant evidence against any defendant in a criminal case," is simply incredible. Surely, this Court's commitment to ensuring that death sentences are imposed reliably and reasonably requires that nonprobative and highly prejudicial testimony on the ultimate question of life or death be excluded from a capital sentencing hearing.

Despite its recognition that the testimony at issue was probably wrong and certainly prejudicial, the Court holds this testimony admissible because the Court is

---

"is proffered by a witness bearing the title of 'Doctor,' its impact on the jury is much greater than if it were not masquerading as something it is not."

... In United States v. Addison, the court observed that scientific evidence may "assume a posture of mystic infallibility in the eyes of a jury of laymen." Another court has noted that scientific evidence "is likely to be shrouded with an aura of near infallibility, akin to the ancient oracle of Delphi." ...

"unconvinced ... that the adversary process cannot be trusted to sort out the reliable from the unreliable evidence and opinion about future dangerousness."
One can only wonder how juries are to separate valid from invalid expert opinions when the "experts" themselves are so obviously unable to do so. Indeed, the evidence suggests that juries are not effective at assessing the validity of scientific evidence.

There can be no question that psychiatric predictions of future violence will have an undue effect on the ultimate verdict. Even judges tend to accept psychiatrists' recommendations about a defendant's dangerousness with little regard for cross-examination or other testimony. The American Bar Association has warned repeatedly that sentencing juries are particularly incapable of dealing with information relating to "the likelihood that the defendant will commit other crimes," and similar predictive judgments. Relying on the ABA's conclusion, the joint opinion announcing the judgment in Gregg v. Georgia, recognized that,

"[s]ince the members of a jury will have had little, if any, previous experience in sentencing, they are unlikely to be skilled in dealing with the information they are given."
But the Court in this case, in its haste to praise the jury's ability to find the truth, apparently forgets this well-known and worrisome shortcoming.

As if to suggest that petitioner's position that unreliable expert testimony should be excluded is unheard of in the law, the Court relies on the proposition that the rules of evidence generally

"anticipate that relevant, unprivileged evidence should be admitted and its weight left to the factfinder, who would have the benefit of cross-examination and contrary evidence by the opposing party."
But the Court simply ignores hornbook law that, despite the availability of cross-examination and rebuttal witnesses,

"opinion evidence is not admissible if the court believes that the state of the pertinent art or scientific knowledge does not permit a reasonable opinion to be asserted."
Because it is feared that the jury will overestimate its probative value, polygraph evidence, for example, almost invariably is excluded from trials despite the fact that, at a conservative estimate, an experienced polygraph examiner can detect truth or deception correctly about 80 to 90 percent of the time. In no area is purportedly "expert" testimony admitted for the jury's consideration where it cannot be demonstrated that it is correct more often than not. "It is inconceivable that a judgment could be considered an expert' judgment when it is less accurate than the flip of a coin." The risk that a jury will be incapable of separating "scientific" myth from reality is deemed unacceptably high.[4]

The Constitution's mandate of reliability, with the stakes at life or death, precludes

[4]The Court observes that this well-established rule is a matter of evidence law, not constitutional law. ... But the principle requiring that capital sentencing procedures ensure reliable verdicts, which the Court ignores, and the principle that due process is violated by the introduction of certain types of seemingly conclusive, but actually unreliable, evidence, ... which the Court also ignores, are constitutional doctrines of long standing. The teaching of the evidence doctrine is that unreliable scientific testimony creates a serious and unjustifiable risk of an erroneous verdict, and that the adversary process, at its best, does not remove this risk. We should not dismiss this lesson merely by labeling the doctrine nonconstitutional; its relevance to the constitutional question before the Court could not be more certain.

reliance on cross-examination and the opportunity to present rebuttal witnesses as an antidote for this distortion of the truthfinding process. Cross-examination is unlikely to reveal the fatuousness of psychiatric predictions because such predictions often rest, as was the case here, on psychiatric categories and intuitive clinical judgments not susceptible to cross-examination and rebuttal. Psychiatric categories have little or no demonstrated relationship to violence, and their use often obscures the unimpressive statistical or intuitive bases for prediction.[5] The APA particularly condemns the use of the diagnosis employed by Doctors Grigson and Holbrook in this case, that of sociopathy:

"In this area confusion reigns. The psychiatrist who is not careful can mislead the judge or jury into believing that a person has a major mental disease simply on the basis of a description of prior criminal behavior. Or a psychiatrist can mislead the court into believing that an individual is devoid of conscience on the basis of a description of criminal acts alone. ... The profession of psychiatry has a responsibility to avoid inflicting this confusion upon the courts, and to spare the defendant the harm that may result. ... Given our uncertainty about the implications of the finding, the diagnosis of sociopathy ... should not be used to justify or to support predictions of future conduct. There is no certainty in this area."

It is extremely unlikely that the adversary process will cut through the facade of superior knowledge. The Chief Justice [Burger] long ago observed:

"The very nature of the adversary system ... complicates the use of scientific opinion evidence, particularly in the field of psychiatry. This system of partisan contention, of attack and counterattack, at its best is not ideally suited to developing an accurate portrait or profile of the human personality, especially in the area of abnormal behavior. Although under ideal conditions the adversary system can develop for a jury most of the necessary fact material for an adequate decision, such conditions are rarely achieved in the courtrooms in this country. These ideal conditions would include a highly skilled and experienced trial judge and highly skilled lawyers on both sides of the case, all of whom, in addition to being well-trained in the law and in the techniques of advocacy, would be sophisticated in matters of medicine, psychiatry, and psychology. It is far too rare that all three of the legal actors in the cast meet these standards."

Another commentator has noted:

"Competent cross-examination and jury instructions may be partial antidotes ... but they cannot be complete. Many of the cases are not truly adversarial; too few attorneys are skilled at cross-examining psychiatrists, laypersons overweigh the testimony of experts, and, in any case, unrestricted use of experts promotes the incorrect view that the questions are primarily scientific. There is, however, no antidote for the major difficulty with mental health 'experts'—that they simply are not experts. ... In realms beyond their true expertise, the law has little special to learn from them; too often, their testimony is ... prejudicial."

Nor is the presentation of psychiatric witnesses on behalf of the defense likely to remove

[5]In one study, for example, the only factor statistically related to whether psychiatrists predicted that a subject would be violent in the future was the type of crime with which the subject was charged. Yet the defendant's charge was mentioned by the psychiatrists to justify their predictions in only one-third of the cases. The criterion most frequently cited was "delusional or impaired thinking." ...

the prejudicial taint of misleading testimony by prosecution psychiatrists. No reputable expert would be able to predict with confidence that the defendant will not be violent; at best, the witness will be able to give his opinion that all predictions of dangerousness are unreliable. Consequently, the jury will not be presented with the traditional battle of experts with opposing views on the ultimate question. Given a choice between an expert who says that he can predict with certainty that the defendant, whether confined in prison or free in society, will kill again, and an expert who says merely that no such prediction can be made, members of the jury, charged by law with making the prediction, surely will be tempted to opt for the expert who claims he can help them in performing their duty, and who predicts dire consequences if the defendant is not put to death.[6]

Moreover, even at best, the presentation of defense psychiatrists will convert the death sentence hearing into a battle of experts, with the Eighth Amendment's well-established requirement of individually focused sentencing a certain loser. The jury's attention inevitably will turn from an assessment of the propriety of sentencing to death the defendant before it to resolving a scientific dispute about the capabilities of psychiatrists to predict future violence. In such an atmosphere, there is every reason to believe that the jury may be distracted from its constitutional responsibility to consider "particularized mitigating factors," in passing on the defendant's future dangerousness.

One searches the Court's opinion in vain for a plausible justification for tolerating the State's creation of this risk of an erroneous death verdict. As one Court of Appeals has observed:

"A courtroom is not a research laboratory. The fate of a defendant ... should not hang on his ability to successfully rebut scientific evidence which bears an 'aura of special reliability and trustworthiness,' although, in reality, the witness is testifying on the basis of an unproved hypothesis ... which has yet to gain general acceptance in its field."

Ultimately, when the Court knows full well that psychiatrists' predictions of dangerousness are specious, there can be no excuse for imposing on the defendant, on pain of his life, the heavy burden of convincing a jury of laymen of the fraud.[7]

[6] "Although jurors may treat mitigating psychiatric evidence with skepticism, they may credit psychiatric evidence demonstrating aggravation. Especially when jurors' sensibilities are offended by a crime, they may seize upon evidence of dangerousness to justify an enhanced sentence." ... Thus, the danger of jury deference to expert opinions is particularly acute in death penalty cases. Expert testimony of this sort may permit juries to avoid the difficult and emotionally draining personal decisions concerning rational and just punishment. ... Doctor Grigson himself has noted both the superfluousness and the misleading effect of his testimony: "I think you could do away with the psychiatrist in these cases. Just take any man off the street, show him what the guy's done, and most of these things are so clear-cut he would say the same things I do. But I think the jurors feel a little better when a psychiatrist says it—somebody that's supposed to know more than they know." ...

[7] The Court is far wide of the mark in asserting that excluding psychiatric predictions of future dangerousness from capital sentencing proceedings "would immediately call into question those other contexts in which predictions of future behavior are constantly made." ... Short-term predictions of future violence, for the purpose of emergency commitment or treatment, are considerably more accurate than long-term predictions. In other contexts where psychiatric predictions of future dangerousness are made, moreover, the subject will not be criminally convicted, much less put to death, as a result of predictive error. The risk of

The Court is simply wrong in claiming that psychiatric testimony respecting future dangerousness is necessarily admissible in light of Jurek v. Texas, or Estelle v. Smith. As the Court recognizes, Jurek involved "only lay testimony." Thus, it is not surprising that "there was no suggestion by the Court that the testimony of doctors would be inadmissible," and it is simply irrelevant that the Jurek Court did not "disapprov[e]" the use of such testimony. In Smith, the psychiatric testimony at issue was given by the same Doctor Grigson who confronts us in this case, and his conclusions were disturbingly similar to those he rendered here. The APA, appearing as *amicus curiae*, argued that all psychiatric predictions of future dangerousness should be excluded from capital sentencing proceedings. The Court did not reach this issue, because it found Smith's death sentence invalid on narrower grounds: Doctor Grigson's testimony had violated Smith's Fifth and Sixth Amendment right. Contrary to the Court's inexplicable assertion in this case, Smith certainly did not reject the APA's position. Rather, the Court made clear that "the holding in Jurek was guided by recognition that the inquiry [into dangerousness] mandated by Texas law does not require resort to medical experts." If Jurek and Smith held that psychiatric predictions of future dangerousness are admissible in a capital sentencing proceeding as the Court claims, this guiding recognition would have been irrelevant.

The Court also errs in suggesting that the exclusion of psychiatrists' predictions of future dangerousness would be contrary to the logic of Jurek. Jurek merely upheld Texas' substantive decision to condition the death sentence upon proof of a probability that the defendant will commit criminal acts of violence in the future. Whether the evidence offered by the prosecution to prove that probability is so unreliable as to violate a capital defendant's rights to due process is an entirely different matter, one raising only questions of fair procedure.[8] Jurek's conclusion that Texas may impose the death penalty on capital defendants who probably will commit criminal acts of violence in no way establishes that the prosecution may convince a jury that this is so by misleading or patently unreliable evidence.

Moreover, Jurek's holding that the Texas death statute is not impermissibly vague does not lead ineluctably to the conclusion that psychiatric testimony is admissible. It makes sense to exclude psychiatric predictions of future violence while admitting lay testimony, because psychiatric predictions appear to come from trained mental health professionals, who purport to have special expertise. In view of the total scientific groundlessness of these predictions,

---

error therefore may be shifted to the defendant to some extent. ... The APA, discussing civil commitment proceedings based on determinations of dangerousness, states that, in light of the unreliability of psychiatric predictions, "[c]lose monitoring, frequent follow-up, and a willingness to change one's mind about treatment recommendations and dispositions for violent persons, whether within the legal system or without, is the only acceptable practice if the psychiatrist is to play a helpful role in these assessments of dangerousness." ... In a capital case, there will be no chance for "follow-up" or "monitoring." A subsequent change of mind brings not justice delayed, but the despair of irreversible error. ...

[8]The Court's focus in the death penalty cases has been primarily on ensuring a fair procedure: "In ensuring that the death penalty is not meted out arbitrarily or capriciously, the Court's principal concern has been more with the procedure by which the State imposes the death sentence than with the substantive factors the State lays before the jury as a basis for imposing death, once it has been determined that the defendant falls within the category of persons eligible for the death penalty."

psychiatric testimony is fatally misleading. Lay testimony, frankly based on statistical factors with demonstrated correlations to violent behavior, would not raise this substantial threat of unreliable and capricious sentencing decisions, inimical to the constitutional standards established in our cases; and such predictions are as accurate as any a psychiatrist could make. Indeed, the very basis of Jurek, as I understood it, was that such judgments can be made by laymen on the basis of lay testimony.

Our constitutional duty is to ensure that the State proves future dangerousness, if at all, in a reliable manner, one that ensures that "any decision to impose the death sentence be, and appear to be, based on reason rather than caprice or emotion." Texas' choice of substantive factors does not justify loading the factfinding process against the defendant through the presentation of what is, at bottom, false testimony.

# Chapter 7

# The Basic Sampling Model and Associated Topics

# Chapter 8

# Psychometrics

## 8.1 Suggested Reading on Psychometric Issues

There are seven items in the Suggested Reading that involve psychometrics in some way; one article is on testing women in science; three are on personality/clinical assessment; two concern cheating on school assessments; and one discusses the difficulties in developing a diagnostic test for the elusive concept of attention deficit hyperactivity disorder:

B.8 – Psychometric Issues

B.8.1 – Bias Called Persistent Hurdle for Women in Sciences (Tamar Lewin, *New York Times*, March 21, 2010)

B.8.2 – Annals of Medicine: The Dictionary of Disorder (Alix Spiegel, *New Yorker*, January 3, 2005)

B.8.3 – Personality Plus (Malcom Gladwell, *New Yorker*, September 20, 2004)

B.8.4 – Experts Say Schools Need to Screen for Cheating (Shaila Dewan, *New York Times*, February 13, 2010)

B.8.5 – Seeking an Objective Test for Attention Disorder (Katherine Ellison, *New York Times*, May 31, 2010)

B.8.6 – Under Pressure, Teachers Tamper With Tests (Trip Gabriel, *New York Times*, June 10, 2010)

B.8.7 – Linking Personality to Brain Structure (Sindya N. Bhanoo, *New York Times*, June 28, 2010)

### 8.1.1 Appendix: Excerpts From Brigham's *A Study of American Intelligence*

The question of the differences that may exist between the various races of man, or between various sub-species of the same race, or between political aggregations of men in nationality groups may easily become the subject of the most acrimonious discussion. The anthropologists of France and Germany, shortly after the close of the Franco-Prussian war, fought another national war on a small scale. It is difficult to keep racial hatreds and antipathies

out of the most scholarly investigations in this field. The debate becomes especially bitter when mental traits are discussed. No one can become very indignant on finding his race classified by its skull dimensions, stature, or hair color, but let a person discover the statement that his race is unintelligent or emotionally unstable, and he is immediately ready to do battle.

Until recent years we have had no methods available for measuring mental traits scientifically, so that the literature on race differences consists largely of opinions of students who are very apt to become biased, when, leaving the solid realm of physical measurements, they enter the more intangible field of estimating mental capacity.

Gradually, however, various investigators using more or less refined psychological measurements commenced to assemble a body of data that will some day reach respectable proportions. Since 1910, we have witnessed in this country a remarkable development in methods of intelligence testing, and these methods have been applied to the study of race differences. Scattered investigations report and compare the intelligence scores of children of white, Negro, or Indian parentage, and sometimes the scores of various nationality or nativity groups. The results of these investigations are, however, almost impossible to correlate, for they have been made by different methods, by different measuring scales, on children of a wide variety of chronological ages, and above all, on comparatively small groups of subjects, so that conclusions based on the studies have no high degree of reliability.

For our purposes in this country, the army mental tests give us an opportunity for a national inventory of our own mental capacity, and the mental capacity of those we have invited to live with us. We find reported in Memoir XV of the National Academy of Sciences, the intelligence scores of about 81,000 native born Americans, 12,000 foreign born individuals, and 23,000 Negroes. From the standpoint of the numbers examined, we have here an investigation which, of course, surpasses in reliability all preceding investigations, assembled and correlated, a hundred fold. These army data constitute the first really significant contribution to the study of race differences in mental traits. They give us a scientific basis for our conclusions. (pp. xix–xx)

If the history of the United States could be written in terms of the movements of European peoples to this continent, the first stage represents a Nordic immigration, for New England in Colonial times was populated by an almost pure Nordic type. There followed then a period of Nordic expansion. The next great movement consisted of the migrations of Western European Mediterraneans and Alpines from Ireland and Germany, a movement which started about 1840, and which had practically stopped by 1890. Since there is a considerable proportion of Nordic blood in Ireland and Germany, we should not regard the original Nordic immigration as a movement which stopped suddenly, but merely as having dwindled to two-fifths or one-half of the total racial stock coming here between 1840 and 1890. The third and last great movement consisted of migrations of the Alpine Slav and the Southern European Mediterraneans to this continent, a movement that started about 1890, and which has not yet ceased. Running parallel with the movements of these European peoples, we have the most sinister development in the history of this continent, the importation of the Negro.

The army mental tests enable us to analyze the elements entering into American intelligence. The intelligence test records of the native born, the foreign born, and the Negro are at our disposal. The records deserve the most serious study. But before considering the results of the army tests, a person should be well informed concerning the nature of the tests, and the manner in which they were constructed. (p. xxi)

The army psychological tests included three types of examination:

(1) Group examination alpha, which included eight different sorts of tests, most of which involved the ability to read English.

(2) Group examination beta, which included seven different sorts of tests, none of which involved the ability either to read English or to understand spoken English, the tests consisting of pictures, designs, etc., and being given by instructions in pantomime.

(3) Individual examinations of two types:

(a) Those involving the use of English, the Stanford revision of the Binet-Simon scale and the point scale, and

(b) Those involving no English, consisting of construction puzzles, etc., the instructions being given by gestures, the "performance scale."

When a detachment reported for psychological examination, the first step was that of separating the English speaking and literate from the non-English speaking or illiterate. Those who were both English speaking and literate were given examination alpha. All others were sent to beta. At the close of examination alpha, all men who had made low scores were sent to beta. After examination beta had been given, the examiners tried to recall for individual examinations all men who had made a low score in beta. In the rush of examining it was impossible to recall all men for individual examinations who should have been given special examinations, and some men were graded on alpha who should have been graded on beta, and vice versa, but most men were properly graded by the rough methods in use. In each one of the examinations the range of scores was so great that most men had an opportunity to score.

The great contribution of the committee that first devised the army examining methods and of the men who subsequently developed additional methods in the army consisted of creating and standardizing group examinations alpha and beta. The methods of individual examining were already in existence, the Stanford-Binet scale being an elaboration of Binet's "mental age" scale, and the tests of the performance scale having been more or less completely worked out by other investigators. The task of examining men in large groups was first carried through successfully in the army. Before the war, many psychologists would have scoffed at the notion of examining two or three hundred men at once by giving them booklets containing different sorts of tests, but the large group examinations became matters of daily routine. Group tests have subsequently been tried out in schools and industries with excellent results from the standpoint of test administration. Indeed, when the army alpha examination was given at Ohio State University in October, 1919, practically the entire student body, 6000 in number, was tested by five examiners in eight hours. In the service, it was found that one examiner could control a group of 200 men with ease. The alpha instructions were read by the examiner, and the men ordered to start and stop at the proper time. Examination beta

was more difficult to administer, and was given to smaller groups.

The statistical methods of treating the results of the army tests used in this study are rather intricate, but the principles involved are easily understood. At the outset we must frankly admit that there were minor errors in the three types of examinations given. We can not correct the type of tests that were used, but we can correct the method of scoring them. Most of the difficulties of scoring arise from the fact that different types of measuring scales were used. During the war, the different scales were converted into one general scale of letter grades (A, B, C+, C, C−, D and D−). This method was rough, and although it answered the purposes of the army at the time, it can not be used in any scientific interpretation of the results.

Examination alpha was scored by finding the score on each of the eight tests, adding to get a total, and then converting the total into a letter grade. Beta was similarly scored. It is apparent that some tests in alpha might be more difficult than others, that some tests in beta might be easier than any test in alpha, and that variations might have occurred which it was impossible to predict at the time the examinations were made. Recognizing these facts, then, the army statisticians worked out another method of scoring the results, which eliminates all of these sources of error. This method is known as the combined scale, a theoretical intelligence scale running from 0 to 25, into which the alpha, beta and individual examination scores may be converted, so that we finally have one measurement instead of three.

Psychological measurements involve much more than creating tests and giving tests. After all the results are in, we still have the problem of interpreting the results, and this interpretation is largely a statistical problem. Too much credit can not be given to the staff of the Psychological Division of the Surgeon General's Office, who continued in the service long after the war was over, patiently studying and analyzing the results. The combined scale was very largely the work of two young psychologists, Carl R. Brown and Mark A. May, and their work on this problem ... is without doubt the greatest contribution that has yet been made to the statistical phases of the science of mental measurement.

The theory underlying the combined scale is simply that of regarding each test of alpha and beta as a separate measuring scale. One group of individuals including 1047 men born in English speaking countries, was examined on alpha, re-examined on beta, and if possible, examined again on the Stanford-Binet scale. This group of 1047 cases constituted the basis on which a method of combining the separate tests into a combined scale was empirically evolved.

From now on in the course of our study of the army test records, we must regard alpha and beta as two booklets containing, in all, fifteen different measuring scales of intelligence. (pp. xxi–xxv)

Our study of the army tests of foreign born individuals has pointed at every step to the conclusion that the average intelligence of our immigrants is declining. This deterioration in the intellectual level of immigrants has been found to be due to two causes. The migrations of the Alpine and Mediterranean races have increased to such an extent in the last thirty or forty years that this blood now constitutes 70% or 75% of the total immigration. The

representatives of the Alpine and Mediterranean races in our immigration are intellectually inferior to the representatives of the Nordic race which formerly made up about 50% of our immigration. In addition, we find that we are getting progressively lower and lower types from each nativity group or race. ...

It is also possible to make a picture of the elements now entering into American intelligence. At one extreme we have the distribution of the Nordic race group. At the other extreme we have the American Negro. Between the Nordic and the Negro, but closer to the Negro than to the Nordic, we find the Alpine and Mediterranean types. (p. 197)

Throughout this study all measurements have been made in terms of averages and variability about the average. In interpreting averages, we must never forget that they stand for an entire distribution. Careless thinkers are prone to select one or two striking examples of ability from a particular group, and then rest confidently in the belief that they have overthrown an argument based on the total distribution of ability. The Fourth of July orator can convincingly raise the popular belief in the intellectual level of Poland by shouting the name of Kosciusko from a high platform, but he can not alter the distribution of the intelligence of the Polish immigrant. All countries send men of exceptional ability to America, but the point is that some send fewer than others. (p. 197; p. 202)

According to all evidence available, then, American intelligence is declining, and will proceed with an accelerating rate as the racial admixture becomes more and more extensive. The decline of American intelligence will be more rapid than the decline of the intelligence of European national groups, owing to the presence here of the Negro. These are the plain, if somewhat ugly, facts that our study shows. The deterioration of American intelligence is not inevitable, however, if public action can be aroused to prevent it. There is no reason why legal steps should not be taken which would insure a continuously progressive upward evolution.

The steps that should be taken to preserve or increase our present intellectual capacity must of course be dictated by science and not by political expediency. Immigration should not only be restrictive but highly selective. And the revision of the immigration and naturalization laws will only afford a slight relief from our present difficulty. The really important steps are those looking toward the prevention of the continued propagation of defective strains in the present population. If all immigration were stopped now, the decline of American intelligence would still be inevitable. This is the problem which must be met, and our manner of meeting it will determine the future course of our national life. (p. 210)

### 8.1.2 Appendix: Racial Integrity Act of 1924 (State of Virginia); Loving v. Virginia

Be it enacted by the general assembly of Virginia, that the State registrar of vital statistics may, as soon as practicable after the taking effect of this act, prepare a form whereon the racial composition of any individual, as Caucasian, Negro, Mongolian, American Indian, Asiatic Indian, Malay, or any mixture thereof, or any other non-Caucasic strains, and if there be any mixture, then, the racial composition of the parents and other ancestors, in

so far as ascertainable, so as to show in what generation such mixture occurred, may be certified by such individual, which form shall be known as a registration certificate.

It shall be a felony for any person wilfully or knowingly to make a registration certificate false as to color or race. The wilful making of a false registration or birth certificate shall be punished by confinement in the penitentiary for one year.

No marriage license shall be granted until the clerk or deputy clerk has reasonable assurance that the statements as to color of both man and woman are correct.

If there is reasonable cause to disbelieve that applicants are of pure white race, when that fact is stated, the clerk or deputy clerk shall withhold the granting of the license until satisfactory proof is produced that both applicants are "white persons" as provided for in this act.

The clerk or deputy clerk shall use the same care to assure himself that both applicants are colored, when that fact is claimed.

It shall hereafter be unlawful for any white person in this State to marry any save a white person, or a person with no other admixture of blood than white and American Indian. For the purpose of this act, the term "white person" shall apply only to the person who has no trace whatsoever of any blood other than Caucasian; but persons who have one-sixteenth or less of the blood of the American Indian and have no other non-Caucasic blood shall be deemed to be white persons. All laws heretofore passed and now in effect regarding the intermarriage of white and colored persons shall apply to marriages prohibited by this act.

---

Loving v. Virginia

WARREN, C.J., Opinion of the Court

SUPREME COURT OF THE UNITED STATES

This case presents a constitutional question never addressed by this Court: whether a statutory scheme adopted by the State of Virginia to prevent marriages between persons solely on the basis of racial classifications violates the Equal Protection and Due Process Clauses of the Fourteenth Amendment. For reasons which seem to us to reflect the central meaning of those constitutional commands, we conclude that these statutes cannot stand consistently with the Fourteenth Amendment.

In June, 1958, two residents of Virginia, Mildred Jeter, a Negro woman, and Richard Loving, a white man, were married in the District of Columbia pursuant to its laws. Shortly after their marriage, the Lovings returned to Virginia and established their marital abode in Caroline County. At the October Term, 1958, of the Circuit Court of Caroline County, a grand jury issued an indictment charging the Lovings with violating Virginia's ban on interracial marriages. On January 6, 1959, the Lovings pleaded guilty to the charge, and were sentenced to one year in jail; however, the trial judge suspended the sentence for a period of 25 years on the condition that the Lovings leave the State and not return to Virginia together for 25 years. He stated in an opinion that:

Almighty God created the races white, black, yellow, malay and red, and he placed them on separate continents. And, but for the interference with his arrangement, there would be no cause for such marriage. The fact that he separated the races shows that he did not

intend for the races to mix.

. . . The Supreme Court of Appeals upheld the constitutionality of the anti-miscegenation statutes and, after modifying the sentence, affirmed the convictions. The Lovings appealed this decision, and we noted probable jurisdiction on December 12, 1966.

The two statutes under which appellants were convicted and sentenced are part of a comprehensive statutory scheme aimed at prohibiting and punishing interracial marriages. The Lovings were convicted of violating . . . the Virginia Code:

Leaving State to evade law. – If any white person and colored person shall go out of this State, for the purpose of being married, and with the intention of returning, and be married out of it, and afterwards return to and reside in it, cohabiting as man and wife, they shall be punished . . . and the marriage shall be governed by the same law as if it had been solemnized in this State. The fact of their cohabitation here as man and wife shall be evidence of their marriage.

. . . the penalty for miscegenation, provides:

Punishment for marriage. – If any white person intermarry with a colored person, or any colored person intermarry with a white person, he shall be guilty of a felony and shall be punished by confinement in the penitentiary for not less than one nor more than five years.

Other central provisions in the Virginia statutory scheme . . . automatically voids all marriages between "a white person and a colored person" without any judicial proceeding, . . . define "white persons" and "colored persons and Indians" for purposes of the statutory prohibitions. The Lovings have never disputed in the course of this litigation that Mrs. Loving is a "colored person" or that Mr. Loving is a "white person" within the meanings given those terms by the Virginia statutes.

Virginia is now one of 16 States which prohibit and punish marriages on the basis of racial classifications. Penalties for miscegenation arose as an incident to slavery, and have been common in Virginia since the colonial period. The present statutory scheme dates from the adoption of the Racial Integrity Act of 1924, passed during the period of extreme nativism which followed the end of the First World War. The central features of this Act, and current Virginia law, are the absolute prohibition of a "white person" marrying other than another "white person," a prohibition against issuing marriage licenses until the issuing official is satisfied that the applicants' statements as to their race are correct, certificates of "racial composition" to be kept by both local and state registrars, and the carrying forward of earlier prohibitions against racial intermarriage.

There can be no question but that Virginia's miscegenation statutes rest solely upon distinctions drawn according to race. The statutes proscribe generally accepted conduct if engaged in by members of different races. Over the years, this Court has consistently repudiated "distinctions between citizens solely because of their ancestry" as being "odious to a free people whose institutions are founded upon the doctrine of equality." At the very least, the Equal Protection Clause demands that racial classifications, especially suspect in criminal statutes, be subjected to the "most rigid scrutiny," and, if they are ever to be upheld, they must be shown to be necessary to the accomplishment of some permissible state objective, independent of the racial discrimination which it was the object of the Fourteenth

Amendment to eliminate. Indeed, two members of this Court have already stated that they cannot conceive of a valid legislative purpose ... which makes the color of a person's skin the test of whether his conduct is a criminal offense.

There is patently no legitimate overriding purpose independent of invidious racial discrimination which justifies this classification. The fact that Virginia prohibits only interracial marriages involving white persons demonstrates that the racial classifications must stand on their own justification, as measures designed to maintain White Supremacy. We have consistently denied the constitutionality of measures which restrict the rights of citizens on account of race. There can be no doubt that restricting the freedom to marry solely because of racial classifications violates the central meaning of the Equal Protection Clause.

These statutes also deprive the Lovings of liberty without due process of law in violation of the Due Process Clause of the Fourteenth Amendment. The freedom to marry has long been recognized as one of the vital personal rights essential to the orderly pursuit of happiness by free men.

Marriage is one of the "basic civil rights of man," fundamental to our very existence and survival. To deny this fundamental freedom on so unsupportable a basis as the racial classifications embodied in these statutes, classifications so directly subversive of the principle of equality at the heart of the Fourteenth Amendment, is surely to deprive all the State's citizens of liberty without due process of law. The Fourteenth Amendment requires that the freedom of choice to marry not be restricted by invidious racial discriminations. Under our Constitution, the freedom to marry, or not marry, a person of another race resides with the individual, and cannot be infringed by the State.

These convictions must be reversed.

It is so ordered.

# Chapter 9

# Background: Data Presentation and Interpretation

## 9.1 Suggested Reading on Data Presentation and Interpretation

Throughout this part of the monograph, various readings will be noted that are listed in the Suggested Reading. In addition to these, there are three broad categories of additional readings that pertain generally to how data are presented, interpreted, and then acted upon; these are within the areas of climate change, a clinical/psychiatric context, and PowerPoint usage:

C.1 – Climate Change Controversy

C.1.1 – IPCC's Himalayan Glacier 'Mistake' Not an Accident (Janet Raloff, *Science News*, January 24, 2010)

C.1.2 – Skeptics Find Fault With U.N. Climate Panel (Elisabeth Rosenthal, *New York Times*, February 9, 2010)

C.1.3 – Climate-Change Debate Is Heating Up in Deep Freeze (John M. Broder, *New York Times*, February 11, 2010)

C.1.4 – Global Warming and Weather Psychology (The Editors, *New York Times*, February 11, 2010)

C.1.5 – Global Weirding Is Here (Thomas L. Friedman, *New York Times*, February 17, 2010)

C.1.6 – Climate Science: Credibility at Risk, Scientists Say (Janet Raloff, *Science News*, February 21, 2010)

C.1.7 – Climate Change (Editorial, *New York Times*, February 22, 2010)

C.1.8 – Climate-Change Fervor Cools Amid Disputed Science (Kim Chipman, *Bloomberg Businessweek*, February 22, 2010)

C.1.9 – Independent Board to Review Work of Top Climate Panel (Reuters, February 27, 2010)

C.1.10 – We Can't Wish Away Climate Change (Al Gore, *New York Times*, February 28, 2010)

C.1.11 – IPCC Looks to Vet, Report Climate-Science Better (Janet Raloff, *Science News*, February 28, 2010)

C.1.12 – Scientists Taking Steps to Defend Work on Climate (John Broder, *New York Times*, March 2, 2010)

C.1.13 – Darwin Foes Add Warming to Targets (Leslie Kaufman, *New York Times*, March 3, 2010)

C.1.14 – Panel Will Review U.N. Climate Work (John M. Broder, *New York Times*, March 10, 2010)

C.1.15 – Among Weathercasters, Doubt on Warming (Leslie Kaufman, *New York Times*, March 29, 2010)

C.1.16 – Climate Fears Turn to Doubts Among Britons (Elisabeth Rosenthal, *New York Times*, May 24, 2010)

C.2 – Clinical and Psychiatric Issues

C.2.1 – Head Case: Can Psychiatry Be a Science? (Louis Menand, *New Yorker*, March 1, 2010)

C.2.2 – Talking Back to Prozac (Frederick C. Crews, *New York Review of Books*, December 6, 2007)

The Truth About Prozac: An Exchange (*New York Review of Books*, February 14, 2008)

Prozac and Sexual Desire (*New York Review of Books*, March 20, 2008)

Branded By Pharma (*New York Review of Books*, May 1, 2008)

C.2.3 – Concocting a Cure for Kids With Issues (Judith Warner, *New York Times*, March 10, 2010)

C.2.4 – Mind Over Meds (Daniel Carlat, *New York Times*, April 19, 2010)

C.2.5 – On the Verge of 'Vital Exhaustion'? (Benedict Carey, *New York Times*, May 31, 2010)

C.3 – Presentation

C.3.1 – We Have Met the Enemy and He Is PowerPoint (Elisabeth Bumiller, *New York Times*, April 26, 2010)

C.3.2 – Essay: Dumb-Dumb Bullets (T.X. Hammes, *Armed Forces Journal*, July, 2009)

### 9.1.1 Appendix: Brown v. Board of Education (Decided: May 17, 1954)

MR. CHIEF JUSTICE WARREN delivered the opinion of the Court.

These cases come to us from the States of Kansas, South Carolina, Virginia, and Delaware. They are premised on different facts and different local conditions, but a common legal question justifies their consideration together in this consolidated opinion. ... In each of the cases, minors of the Negro race, through their legal representatives, seek the aid of the courts in obtaining admission to the public schools of their community on a nonsegregated basis. In each instance, ... they had been denied admission to schools attended by white

children under laws requiring or permitting segregation according to race. This segregation was alleged to deprive the plaintiffs of the equal protection of the laws under the Fourteenth Amendment. In each of the cases other than the Delaware case, a three-judge federal district court denied relief to the plaintiffs on the so-called "separate but equal" doctrine announced by this Court in Plessy v. Ferguson, ... Under that doctrine, equality of treatment is accorded when the races are provided substantially equal facilities, even though these facilities be separate. In the Delaware case, the Supreme Court of Delaware adhered to that doctrine, but ordered that the plaintiffs be admitted to the white schools because of their superiority to the Negro schools.

The plaintiffs contend that segregated public schools are not "equal" and cannot be made "equal," and that hence they are deprived of the equal protection of the laws. Because of the obvious importance of the question presented, the Court took jurisdiction. ...

Reargument was largely devoted to the circumstances surrounding the adoption of the Fourteenth Amendment in 1868. It covered exhaustively consideration of the Amendment in Congress, ratification by the states, then existing practices in racial segregation, and the views of proponents and opponents of the Amendment. This discussion and our own investigation convince us that, although these sources cast some light, it is not enough to resolve the problem with which we are faced. At best, they are inconclusive. The most avid proponents of the post-War Amendments undoubtedly intended them to remove all legal distinctions among "all persons born or naturalized in the United States." Their opponents, just as certainly, were antagonistic to both the letter and the spirit of the Amendments and wished them to have the most limited effect. What others in Congress and the state legislatures had in mind cannot be determined with any degree of certainty.

An additional reason for the inconclusive nature of the Amendment's history, with respect to segregated schools, is the status of public education at that time. In the South, the movement toward free common schools, supported ... by general taxation, had not yet taken hold. Education of white children was largely in the hands of private groups. Education of Negroes was almost nonexistent, and practically all of the race were illiterate. In fact, any education of Negroes was forbidden by law in some states. Today, in contrast, many Negroes have achieved outstanding success in the arts and sciences as well as in the business and professional world. It is true that public school education at the time of the Amendment had advanced further in the North, but the effect of the Amendment on Northern States was generally ignored in the congressional debates. Even in the North, the conditions of public education did not approximate those existing today. The curriculum was usually rudimentary; ungraded schools were common in rural areas; the school term was but three months a year in many states; and compulsory school attendance was virtually unknown. As a consequence, it is not surprising that there should be so little in the history of the Fourteenth Amendment relating to its intended effect on public education.

In the first cases in this Court construing the Fourteenth Amendment, decided shortly after its adoption, the Court interpreted it as proscribing all state-imposed discriminations against the Negro race. The doctrine of ... "separate but equal" did not make its appearance in this Court until 1896 in the case of Plessy v. Ferguson, ... involving not education but

transportation. American courts have since labored with the doctrine for over half a century. In this Court, there have been six cases involving the "separate but equal" doctrine in the field of public education. ... the validity of the doctrine itself was not challenged. In more recent cases, all on the graduate school ... level, inequality was found in that specific benefits enjoyed by white students were denied to Negro students of the same educational qualifications. ... In none of these cases was it necessary to re-examine the doctrine to grant relief to the Negro plaintiff. And in Sweatt v. Painter, ... the Court expressly reserved decision on the question whether Plessy v. Ferguson should be held inapplicable to public education.

In the instant cases, that question is directly presented. Here, unlike Sweatt v. Painter, there are findings below that the Negro and white schools involved have been equalized, or are being equalized, with respect to buildings, curricula, qualifications and salaries of teachers, and other "tangible" factors. Our decision, therefore, cannot turn on merely a comparison of these tangible factors in the Negro and white schools involved in each of the cases. We must look instead to the effect of segregation itself on public education.

In approaching this problem, we cannot turn the clock back to 1868 when the Amendment was adopted, or even to 1896 when Plessy v. Ferguson was written. We must consider public education in the light of its full development and its present place in American life throughout ... the Nation. Only in this way can it be determined if segregation in public schools deprives these plaintiffs of the equal protection of the laws.

Today, education is perhaps the most important function of state and local governments. Compulsory school attendance laws and the great expenditures for education both demonstrate our recognition of the importance of education to our democratic society. It is required in the performance of our most basic public responsibilities, even service in the armed forces. It is the very foundation of good citizenship. Today it is a principal instrument in awakening the child to cultural values, in preparing him for later professional training, and in helping him to adjust normally to his environment. In these days, it is doubtful that any child may reasonably be expected to succeed in life if he is denied the opportunity of an education. Such an opportunity, where the state has undertaken to provide it, is a right which must be made available to all on equal terms.

We come then to the question presented: Does segregation of children in public schools solely on the basis of race, even though the physical facilities and other "tangible" factors may be equal, deprive the children of the minority group of equal educational opportunities? We believe that it does.

In Sweatt v. Painter, ... in finding that a segregated law school for Negroes could not provide them equal educational opportunities, this Court relied in large part on "those qualities which are incapable of objective measurement but which make for greatness in a law school." In McLaurin v. Oklahoma State Regents, ... the Court, in requiring that a Negro admitted to a white graduate school be treated like all other students, again resorted to intangible considerations: " ... his ability to study, to engage in discussions and exchange views with other students, and, in general, to learn his profession." ... Such considerations apply with added force to children in grade and high schools. To separate them from others

of similar age and qualifications solely because of their race generates a feeling of inferiority as to their status in the community that may affect their hearts and minds in a way unlikely ever to be undone. The effect of this separation on their educational opportunities was well stated by a finding in the Kansas case by a court which nevertheless felt compelled to rule against the Negro plaintiffs:

"Segregation of white and colored children in public schools has a detrimental effect upon the colored children. The impact is greater when it has the sanction of the law; for the policy of separating the races is usually interpreted as denoting the inferiority of the negro group. A sense of inferiority affects the motivation of a child to learn. Segregation with the sanction of law, therefore, has a tendency to [retard] the educational and mental development of negro children and to deprive them of some of the benefits they would receive in a racial[ly] integrated school system."

Whatever may have been the extent of psychological knowledge at the time of Plessy v. Ferguson, this finding is amply supported by modern authority. Any language ... in Plessy v. Ferguson contrary to this finding is rejected. We conclude that in the field of public education the doctrine of "separate but equal" has no place. Separate educational facilities are inherently unequal. Therefore, we hold that the plaintiffs and others similarly situated for whom the actions have been brought are, by reason of the segregation complained of, deprived of the equal protection of the laws guaranteed by the Fourteenth Amendment. This disposition makes unnecessary any discussion whether such segregation also violates the Due Process Clause of the Fourteenth Amendment.

Because these are class actions, because of the wide applicability of this decision, and because of the great variety of local conditions, the formulation of decrees in these cases presents problems of considerable complexity. On reargument, the consideration of appropriate relief was necessarily subordinated to the primary question – the constitutionality of segregation in public education. We have now announced that such segregation is a denial of the equal protection of the laws. ... It is so ordered.

### 9.1.2 Appendix: Matrixx Initiatives, Inc. v. Siracusano (Decided: March 22, 2011)

It might be helpful to define a few of the more esoteric legal terms used:

To act with *scienter* implies a mental state embracing the intent to deceive, manipulate, or defraud;

A *reasonable-person standard* refers to a hypothetical individual exercising average care, skill, and judgement in conduct; a reasonable person acts as a comparative standard for determining liability;

Something is *dispositive* when it is decisive or conclusive, and, for example, settles a dispute or question;

When arguing the case of either party, one is said to *plead* the case;

Something has *materiality* if it is relevant and consequential to an issue under discussion. It does not have to be "statistically significant" to be material. In an earlier court ruling,

the Supreme Court stated: "an omitted fact is material if there is a substantial likelihood that a reasonable shareholder would consider it important in deciding how to vote";

A *bright-line* rule is an absolute criterion ("Bright-line rule," 2011); here, statistical significance is *not* such a bright-line rule when deciding about evidence disclosure to investors buying Matrixx stock.

The Securities Exchange Act of 1934, Section 10: Manipulative and Deceptive Devices, states:
It shall be unlawful for any person, directly or indirectly, by the use of any means or instrumentality of interstate commerce or of the mails, or of any facility of any national securities exchange –

... [and in section 10(b)]

b. To use or employ, in connection with the purchase or sale of any security registered on a national securities exchange or any security not so registered, or any securities-based swap agreement (as defined in section 206B of the Gramm-Leach-Bliley Act), any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the protection of investors.

The Securities Exchange Commission Rule 10b-5, Employment of Manipulative and Deceptive Practices:
It shall be unlawful for any person, directly or indirectly, by the use of any means or instrumentality of interstate commerce, or of the mails or of any facility of any national securities exchange,

(a) To employ any device, scheme, or artifice to defraud,

(b) To make any untrue statement of a material fact or to omit to state a material fact necessary ... to make the statements made, in the light of the circumstances under which they were made, not misleading, or

(c) To engage in any act, practice, or course of business which operates or would operate as a fraud or deceit upon any person, in connection with the purchase or sale of any security.

SUPREME COURT OF THE UNITED STATES
ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT

JUSTICE SOTOMAYOR delivered the opinion of the Court.

This case presents the question whether a plaintiff can state a claim for securities fraud under [section] 10(b) of the Securities Exchange Act of 1934, ... and Securities and Exchange Commission (SEC) Rule 10b-5, ... based on a pharmaceutical company's failure to disclose reports of adverse events associated with a product if the reports do not disclose a statistically significant number of adverse events. Respondents, plaintiffs in a securities fraud class action, allege that petitioners, Matrixx Initiatives, Inc., and three of its executives (collectively Matrixx), failed to disclose reports of a possible link between its leading product, a cold remedy, and loss of smell, rendering statements made by Matrixx misleading. Matrixx contends that respondents' complaint does not adequately allege that Matrixx made

a material representation or omission or that it acted with scienter because the complaint does not allege that Matrixx knew of a statistically significant number of adverse events requiring disclosure. We conclude that the materiality of adverse event reports cannot be reduced to a bright-line rule. Although in many cases reasonable investors would not consider reports of adverse events to be material information, respondents have alleged facts plausibly suggesting that reasonable investors would have viewed these particular reports as material. Respondents have also alleged facts "giving rise to a strong inference" that Matrixx "acted with the required state of mind." ... We therefore hold, in agreement with the Court of Appeals for the Ninth Circuit, that respondents have stated a claim under [section] 10(b) and Rule 10b-5.

...

Through a wholly owned subsidiary, Matrixx develops, manufactures, and markets over-the-counter pharmaceutical products. Its core brand of products is called Zicam. All of the products sold under the name Zicam are used to treat the common cold and associated symptoms. At the time of the events in question, one of Matrixx's products was Zicam Cold Remedy, which came in several forms including nasal spray and gel. The active ingredient in Zicam Cold Remedy was zinc gluconate. Respondents allege that Zicam Cold Remedy accounted for approximately 70 percent of Matrixx's sales.

Respondents initiated this securities fraud class action against Matrixx on behalf of individuals who purchased Matrixx securities between October 22, 2003, and February 6, 2004. The action principally arises out of statements that Matrixx made during the class period relating to revenues and product safety. Respondents claim that Matrixx's statements were misleading in light of reports that Matrixx had received, but did not disclose, about consumers who had lost their sense of smell (a condition called anosmia) after using Zicam Cold Remedy. Respondents' consolidated amended complaint alleges the following facts, which the courts below properly assumed to be true. ...

In 1999, Dr. Alan Hirsch, neurological director of the Smell & Taste Treatment and Research Foundation, Ltd., called Matrixx's customer service line after discovering a possible link between Zicam nasal gel and a loss of smell "in a cluster of his patients." ... Dr. Hirsch told a Matrixx employee that "previous studies had demonstrated that intranasal application of zinc could be problematic." ... He also told the employee about at least one of his patients who did not have a cold and who developed anosmia after using Zicam.

In September 2002, Timothy Clarot, Matrixx's vice president for research and development, called Miriam Linschoten, Ph.D., at the University of Colorado Health Sciences Center after receiving a complaint from a person Linschoten was treating who had lost her sense of smell after using Zicam. Clarot informed Linschoten that Matrixx had received similar complaints from other customers. Linschoten drew Clarot's attention to "previous studies linking zinc sulfate to loss of smell." ... Clarot gave her the impression that he had not heard of the studies. She asked Clarot whether Matrixx had done any studies of its own; he responded that it had not but that it had hired a consultant to review the product. Soon thereafter, Linschoten sent Clarot abstracts of the studies she had mentioned. Research from the 1930's and 1980's had confirmed "[z]inc's toxicity." ... Clarot called Linschoten to ask

whether she would be willing to participate in animal studies that Matrixx was planning, but she declined because her focus was human research.

By September 2003, one of Linschoten's colleagues at the University of Colorado, Dr. Bruce Jafek, had observed 10 patients suffering from anosmia after Zicam use. Linschoten and Jafek planned to present their findings at a meeting of the American Rhinologic Society in a poster presentation entitled "Zicam Induced Anosmia." ... The American Rhinologic Society posted their abstract in advance of the meeting. The presentation described in detail a 55-year old man with previously normal taste and smell who experienced severe burning in his nose, followed immediately by a loss of smell, after using Zicam. It also reported 10 other Zicam users with similar symptoms.

Matrixx learned of the doctors' planned presentation. Clarot sent a letter to Dr. Jafek warning him that he did not have permission to use Matrixx's name or the names of its products. Dr. Jafek deleted the references to Zicam in the poster before presenting it to the American Rhinologic Society.

The following month, two plaintiffs commenced a product liability lawsuit against Matrixx alleging that Zicam had damaged their sense of smell. By the end of the class period on February 6, 2004, nine plaintiffs had filed four lawsuits.

Respondents allege that Matrixx made a series of public statements that were misleading in light of the foregoing information. In October 2003, after they had learned of Dr. Jafek's study and after Dr. Jafek had presented his findings to the American Rhinologic Society, Matrixx stated that Zicam was "'poised for growth in the upcoming cough and cold season'" and that the company had "'very strong momentum.'" ... Matrixx further expressed its expectation that revenues would "'be up in excess of 50% and that earnings, per share for the full year [would] be in the 25 to 30 cent range.'" ... In January 2004, Matrixx raised its revenue guidance, predicting an increase in revenues of 80 percent and earnings per share in the 33-to-38-cent range.

In its Form 10Q filed with the SEC in November 2003, Zicam warned of the potential "'material adverse effect'" that could result from product liability claims, "'whether or not proven to be valid.'" ... It stated that product liability actions could materially affect Matrixx's "'product branding and goodwill,'" leading to reduced customer acceptance. ... It did not disclose, however, that two plaintiffs had already sued Matrixx for allegedly causing them to lose their sense of smell.

On January 30, 2004, Dow Jones News wires reported that the Food and Drug Administration (FDA) was "'looking into complaints that an over-the-counter common-cold medicine manufactured by a unit of Matrixx Initiatives, Inc. ... may be causing some users to lose their sense of smell'" in light of at least three product liability lawsuits. ... Matrixx's stock fell from $13.55 to $11.97 per share after the report. In response, on February 2, Matrixx issued a press release that stated:

"All Zicam products are manufactured and marketed according to FDA guidelines for homeopathic medicine. Our primary concern is the health and safety of our customers and the distribution of factual information about our products. Matrixx believes statements alleging that intranasal Zicam products caused anosmia (loss of smell) are completely un-

founded and misleading.

In no clinical trial of intranasal zinc gluconate gel products has there been a single report of lost or diminished olfactory function (sense of smell). Rather, the safety and efficacy of zinc gluconate for the treatment of symptoms related to the common cold have been well established in two double-blind, placebo controlled, randomized clinical trials. In fact, in neither study were there any reports of anosmia related to the use of this compound. The overall incidence of adverse events associated with zinc gluconate was extremely low, with no statistically significant difference between the adverse event rates for the treated and placebo subsets.

A multitude of environmental and biologic influences are known to affect the sense of smell. Chief among them is the common cold. As a result, the population most likely to use cold remedy products is already at increased risk of developing anosmia. Other common causes of olfactory dysfunction include age, nasal and sinus infections, head trauma, anatomical obstructions, and environmental irritants."

The day after Matrixx issued this press release, its stock price bounced back to $13.40 per share. On February 6, 2004, the end of the class period, Good Morning America, a nationally broadcast morning news program, highlighted Dr. Jafek's findings. ... The program reported that Dr. Jafek had discovered more than a dozen patients suffering from anosmia after using Zicam. It also noted that four lawsuits had been filed against Matrixx. The price of Matrixx stock plummeted to $9.94 per share that same day. Zicam again issued a press release largely repeating its February 2 statement.

On February 19, 2004, Matrixx filed a Form 8-K with the SEC stating that it had "'convened a two-day meeting of physicians and scientists to review current information on smell disorders'" in response to Dr. Jafek's presentation. ... According to the Form 8-K, "'In the opinion of the panel, there is insufficient scientific evidence at this time to determine if zinc gluconate, when used as recommended, affects a person's ability to smell.'" ... A few weeks later, a reporter quoted Matrixx as stating that it would begin conducting "'animal and human studies to further characterize these post-marketing complaints.'" ...

On the basis of these allegations, respondents claimed that Matrixx violated [section] 10(b) of the Securities Exchange Act and SEC Rule 10b-5 by making untrue statements of fact and failing to disclose material facts necessary to make the statements not misleading in an effort to maintain artificially high prices for Matrixx securities.

... Matrixx moved to dismiss respondents' complaint, arguing that they had failed to plead the elements of a material misstatement or omission and scienter. The District Court granted the motion to dismiss. Relying on In re Carter-Wallace, Inc., Securities Litigation, ... it held that respondents had not alleged a "statistically significant correlation between the use of Zicam and anosmia so as to make failure to public[ly] disclose complaints and the University of Colorado study a material omission." ... The District Court similarly agreed that respondents had not stated with particularity facts giving rise to a strong inference of scienter. ... It noted that the complaint failed to allege that Matrixx disbelieved its statements about Zicam's safety or that any of the defendants profited or attempted to profit from Matrixx's public statements. ...

The Court of Appeals reversed. ... Noting that "'[t]he determination [of materiality] requires delicate assessments of the inferences a 'reasonable shareholder' would draw from a given set of facts and the significance of those inferences to him,'" ... the Court of Appeals held that the District Court had erred in requiring an allegation of statistical significance to establish materiality. It concluded, to the contrary, that the complaint adequately alleged "information regarding the possible link between Zicam and anosmia" that would have been significant to a reasonable investor. ... Turning to scienter, the Court of Appeals concluded that "[w]ithholding reports of adverse effects of and lawsuits concerning the product responsible for the company's remarkable sales increase is 'an extreme departure from the standards of ordinary care,'" giving rise to a strong inference of scienter. ...

We granted certiorari, and we now affirm.

...

Section 10(b) of the Securities Exchange Act makes it unlawful for any person to "use or employ, in connection with the purchase or sale of any security ... any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the protection of investors." ... SEC Rule 10b-5 implements this provision by making it unlawful to, among other things, "make any untrue statement of a material fact or to omit to state a material fact necessary ... to make the statements made, in the light of the circumstances under which they were made, not misleading." ... We have implied a private cause of action from the text and purpose of [section] 10(b).

To prevail on their claim that Matrixx made material misrepresentations or omissions in violation of [section] 10(b) and Rule 10b-5, respondents must prove "(1) a material misrepresentation or omission by the defendant; (2) scienter; (3) a connection between the misrepresentation or omission and the purchase or sale of a security; (4) reliance upon the misrepresentation or omission; (5) economic loss; and (6) loss causation." ... Matrixx contends that respondents have failed to plead both the element of a material misrepresentation or omission and the element of scienter because they have not alleged that the reports received by Matrixx reflected statistically significant evidence that Zicam caused anosmia. We disagree.

...

We first consider Matrixx's argument that "adverse event reports that do not reveal a statistically significant increased risk of adverse events from product use are not material information." ...

To prevail on a [section] 10(b) claim, a plaintiff must show that the defendant made a statement that was "misleading as to a material fact." [citing a court case called *Basic*] ... In *Basic*, we held that this materiality requirement is satisfied when there is "'a substantial likelihood that the disclosure of the omitted fact would have been viewed by the reasonable investor as having significantly altered the 'total mix' of information made available.'" ... We were "careful not to set too low a standard of materiality," for fear that management would "'bury the shareholders in an avalanche of trivial information.'" ...

*Basic* involved a claim that the defendant had made misleading statements denying that it

was engaged in merger negotiations when it was, in fact, conducting preliminary negotiations. . . . The defendant urged a bright-line rule that preliminary merger negotiations are material only once the parties to the negotiations reach an agreement in principle. . . . We observed that "[a]ny approach that designates a single fact or occurrence as always determinative of an inherently fact-specific finding such as materiality, must necessarily be overinclusive or underinclusive." . . . We thus rejected the defendant's proposed rule, explaining that it would "artificially exclud[e] from the definition of materiality information concerning merger discussions, which would otherwise be considered significant to the trading decision of a reasonable investor." . . .

Like the defendant in *Basic*, Matrixx urges us to adopt a bright-line rule that reports of adverse events associated with a pharmaceutical company's products cannot be material absent a sufficient number of such reports to establish a statistically significant risk that the product is in fact causing the events. Absent statistical significance, Matrixx argues, adverse event reports provide only "anecdotal" evidence that "the user of a drug experienced an adverse event at some point during or following the use of that drug." . . . Accordingly, it contends, reasonable investors would not consider such reports relevant unless they are statistically significant because only then do they "reflect a scientifically reliable basis for inferring a potential causal link between product use and the adverse event." . . .

As in *Basic*, Matrixx's categorical rule would "artificially exclud[e]" information that "would otherwise be considered significant to the trading decision of a reasonable investor." . . . Matrixx's argument rests on the premise that statistical significance is the only reliable indication of causation. This premise is flawed: As the SEC points out, "medical researchers . . . consider multiple factors in assessing causation." . . . Statistically significant data are not always available. For example, when an adverse event is subtle or rare, "an inability to obtain a data set of appropriate quality or quantity may preclude a finding of statistical significance." . . . Moreover, ethical considerations may prohibit researchers from conducting randomized clinical trials to confirm a suspected causal link for the purpose of obtaining statistically significant data.

. . .

A lack of statistically significant data does not mean that medical experts have no reliable basis for inferring a causal link between a drug and adverse events. As Matrixx itself concedes, medical experts rely on other evidence to establish an inference of causation. . . . We note that courts frequently permit expert testimony on causation based on evidence other than statistical significance. . . . We need not consider whether the expert testimony was properly admitted in those cases, and we do not attempt to define here what constitutes reliable evidence of causation. It suffices to note that, as these courts have recognized, "medical professionals and researchers do not limit the data they consider to the results of randomized clinical trials or to statistically significant evidence." . . .

The FDA similarly does not limit the evidence it considers for purposes of assessing causation and taking regulatory action to statistically significant data. In assessing the safety risk posed by a product, the FDA considers factors such as "strength of the association," "temporal relationship of product use and the event," "consistency of findings across

available data sources," "evidence of a dose-response for the effect," "biologic plausibility," "seriousness of the event relative to the disease being treated," "potential to mitigate the risk in the population," "feasibility of further study using observational or controlled clinical study designs," and "degree of benefit the product provides, including availability of other therapies." ... It "does not apply any single metric for determining when additional inquiry or action is necessary, and it certainly does not insist upon 'statistical significance.'"

Not only does the FDA rely on a wide range of evidence of causation, it sometimes acts on the basis of evidence that suggests, but does not prove, causation. For example, the FDA requires manufacturers of over-the-counter drugs to revise their labeling "to include a warning as soon as there is reasonable evidence of an association of a serious hazard with a drug; a causal relationship need not have been proved." ... More generally, the FDA may make regulatory decisions against drugs based on postmarketing evidence that gives rise to only a suspicion of causation. ... "[A]chieving certain proof of causality through postmarketing surveillance is unusual. Attaining a prominent degree of suspicion is much more likely, and may be considered a sufficient basis for regulatory decisions."

This case proves the point. In 2009, the FDA issued a warning letter to Matrixx stating that "[a] significant and growing body of evidence substantiates that the Zicam Cold Remedy intranasal products may pose a serious risk to consumers who use them." ... The letter cited as evidence 130 reports of anosmia the FDA had received, the fact that the FDA had received few reports of anosmia associated with other intranasal cold remedies, and "evidence in the published scientific literature that various salts of zinc can damage olfactory function in animals and humans." It did not cite statistically significant data.

Given that medical professionals and regulators act on the basis of evidence of causation that is not statistically significant, it stands to reason that in certain cases reasonable investors would as well. As Matrixx acknowledges, adverse event reports "appear in many forms, including direct complaints by users to manufacturers, reports by doctors about reported or observed patient reactions, more detailed case reports published by doctors in medical journals, or larger scale published clinical studies." ... As a result, assessing the materiality of adverse event reports is a "fact-specific" inquiry, ... that requires consideration of the source, content, and context of the reports. This is not to say that statistical significance (or the lack thereof) is irrelevant—only that it is not dispositive of every case.

Application of *Basic's* "total mix" standard does not mean that pharmaceutical manufacturers must disclose all reports of adverse events. Adverse event reports are daily events in the pharmaceutical industry; in 2009, the FDA entered nearly 500,000 such reports into its reporting system, ... The fact that a user of a drug has suffered an adverse event, standing alone, does not mean that the drug caused that event. ... The question remains whether a reasonable investor would have viewed the nondisclosed information "'as having significantly altered the "total mix" of information made available.'" ... For the reasons just stated, the mere existence of reports of adverse events—which says nothing in and of itself about whether the drug is causing the adverse events —will not satisfy this standard. Something more is needed, but that something more is not limited to statistical significance and can come from "the source, content, and context of the reports," ... This contextual

inquiry may reveal in some cases that reasonable investors would have viewed reports of adverse events as material even though the reports did not provide statistically significant evidence of a causal link.

Moreover, it bears emphasis that [section] 10(b) and Rule 10b-5(b) do not create an affirmative duty to disclose any and all material information. Disclosure is required under these provisions only when necessary "to make ... statements made, in the light of the circumstances under which they were made, not misleading ... . Even with respect to information that a reasonable investor might consider material, companies can control what they have to disclose under these provisions by controlling what they say to the market.

... Applying *Basic's* "total mix" standard in this case, we conclude that respondents have adequately pleaded materiality. This is not a case about a handful of anecdotal reports, as Matrixx suggests. Assuming the complaint's allegations to be true, as we must, Matrixx received information that plausibly indicated a reliable causal link between Zicam and anosmia. That information included reports from three medical professionals and researchers about more than 10 patients who had lost their sense of smell after using Zicam. Clarot told Linschoten that Matrixx had received additional reports of anosmia. (In addition, during the class period, nine plaintiffs commenced four product liability lawsuits against Matrixx alleging a causal link between Zicam use and anosmia.) Further, Matrixx knew that Linschoten and Dr. Jafek had presented their findings about a causal link between Zicam and anosmia to a national medical conference devoted to treatment of diseases of the nose. Their presentation described a patient who experienced severe burning in his nose, followed immediately by a loss of smell, after using Zicam—suggesting a temporal relationship between Zicam use and anosmia.

Critically, both Dr. Hirsch and Linschoten had also drawn Matrixx's attention to previous studies that had demonstrated a biological causal link between intranasal application of zinc and anosmia. Before his conversation with Linschoten, Clarot, Matrixx's vice president of research and development, was seemingly unaware of these studies, and the complaint suggests that, as of the class period, Matrixx had not conducted any research of its own relating to anosmia. ... (referencing a press report, issued after the end of the class period, noting that Matrixx said it would begin conducting " 'animal and human studies to further characterize these postmarketing complaints' "). Accordingly, it can reasonably be inferred from the complaint that Matrixx had no basis for rejecting Dr. Jafek's findings out of hand.

We believe that these allegations suffice to "raise a reasonable expectation that discovery will reveal evidence" satisfying the materiality requirement, ... and to "allo[w] the court to draw the reasonable inference that the defendant is liable for the misconduct alleged," ... The information provided to Matrixx by medical experts revealed a plausible causal relationship between Zicam Cold Remedy and anosmia. Consumers likely would have viewed the risk associated with Zicam (possible loss of smell) as substantially outweighing the benefit of using the product (alleviating cold symptoms), particularly in light of the existence of many alternative products on the market. Importantly, Zicam Cold Remedy allegedly accounted for 70 percent of Matrixx's sales. Viewing the allegations of the complaint as a whole, the complaint alleges facts suggesting a significant risk to the commercial viability of Matrixx's

leading product.

It is substantially likely that a reasonable investor would have viewed this information "'as having significantly altered the "total mix" of information made available.'" ... Matrixx told the market that revenues were going to rise 50 and then 80 percent. Assuming the complaint's allegations to be true, however, Matrixx had information indicating a significant risk to its leading revenue-generating product. Matrixx also stated that reports indicating that Zicam caused anosmia were "'completely unfounded and misleading'" and that "'the safety and efficacy of zinc gluconate for the treatment of symptoms related to the common cold have been well established.'" Importantly, however, Matrixx had evidence of a biological link between Zicam's key ingredient and anosmia, and it had not conducted any studies of its own to disprove that link. In fact, as Matrixx later revealed, the scientific evidence at that time was "'insufficient ... to determine if zinc gluconate, when used as recommended, affects a person's ability to smell.'" ...

Assuming the facts to be true, these were material facts "necessary ... to make the statements made, in the light of the circumstances under which they were made, not misleading." ... We therefore affirm the Court of Appeals' holding that respondents adequately pleaded the element of a material misrepresentation or omission.

... Matrixx also argues that respondents failed to allege facts plausibly suggesting that it acted with the required level of scienter. "To establish liability under [section] 10(b) and Rule 10b-5, a private plaintiff must prove that the defendant acted with scienter, 'a mental state embracing intent to deceive, manipulate, or defraud.'" ... We have not decided whether recklessness suffices to fulfill the scienter requirement. ... Because Matrixx does not challenge the Court of Appeals' holding that the scienter requirement may be satisfied by a showing of "deliberate recklessness," ... we assume, without deciding, that the standard applied by the Court of Appeals is sufficient to establish scienter.

Under the PSLRA [Private Securities Litigation Reform Act of 1995], a plaintiff must "state with particularity facts giving rise to a strong inference that the defendant acted with the required state of mind." ... This standard requires courts to take into account "plausible opposing inferences." ... A complaint adequately pleads scienter under the PSLRA "only if a reasonable person would deem the inference of scienter cogent and at least as compelling as any opposing inference one could draw from the facts alleged." ... In making this determination, the court must review "all the allegations holistically." ... The absence of a motive allegation, though relevant, is not dispositive. ...

Matrixx argues, in summary fashion, that because respondents do not allege that it knew of statistically significant evidence of causation, there is no basis to consider the inference that it acted recklessly or knowingly to be at least as compelling as the alternative inferences. "Rather," it argues, "the most obvious inference is that petitioners did not disclose the [reports] simply because petitioners believed they were far too few ... to indicate anything meaningful about adverse reactions to use of Zicam." ... Matrixx's proposed bright-line rule requiring an allegation of statistical significance to establish a strong inference of scienter is just as flawed as its approach to materiality.

The inference that Matrixx acted recklessly (or intentionally, for that matter) is at least

as compelling, if not more compelling, than the inference that it simply thought the reports did not indicate anything meaningful about adverse reactions. According to the complaint, Matrixx was sufficiently concerned about the information it received that it informed Linschoten that it had hired a consultant to review the product, asked Linschoten to participate in animal studies, and convened a panel of physicians and scientists in response to Dr. Jafek's presentation. It successfully prevented Dr. Jafek from using Zicam's name in his presentation on the ground that he needed Matrixx's permission to do so. Most significantly, Matrixx issued a press release that suggested that studies had confirmed that Zicam does not cause anosmia when, in fact, it had not conducted any studies relating to anosmia and the scientific evidence at that time, according to the panel of scientists, was insufficient to determine whether Zicam did or did not cause anosmia. These allegations, "taken collectively," give rise to a "cogent and compelling" inference that Matrixx elected not to disclose the reports of adverse events not because it believed they were meaningless but because it understood their likely effect on the market. ... "[A] reasonable person" would deem the inference that Matrixx acted with deliberate recklessness (or even intent) "at least as compelling as any opposing inference one could draw from the facts alleged." ... We conclude, in agreement with the Court of Appeals, that respondents have adequately pleaded scienter. Whether respondents can ultimately prove their allegations and establish scienter is an altogether different question.

For the reasons stated, the judgment of the Court of Appeals for the Ninth Circuit is Affirmed.

# Chapter 10

# (Mis)reporting of Data

### 10.0.3 Appendix: P. Lorillard Co. v. Federal Trade Commission (Court of Appeals Fourth Circuit; Decided: December 29, 1950)

In *P. Lorillard Co. v. FTC*, the company was charged by the FTC with making a distorted use of a *Reader's Digest* article that discussed the harmful effects of various brands of cigarettes. A laboratory had concluded that no particular brand of cigarettes was substantially more harmful than any other. A table of variations in brand characteristics was inserted in the article to show the insignificance of the differences that existed in the tar and nicotine content of the smoke produced by the various brands. The table indicated that Old Golds had less nicotine and tars, although the difference was so small as to be insignificant. Lorillard launched a national advertising campaign stressing that the *Reader's Digest* test proved that its brand was "lowest in nicotine and tars," and defended its advertising before the FTC on the ground that it had truthfully reported what had been stated in the article. In a 1950 decision, the Fourth Circuit Court of Appeals, upholding the commission's cease-and-desist order, declared that Lorillard's advertising violated the FTC Act because, by printing only a small part of the article, it created an entirely false and misleading impression. "To tell less than the whole truth is a well-known method of deception," the court ruled (Source: Susan Wagner, *Cigarette Country: Tobacco in American History* (1971), pp 72–73).

From the Appeals Court decision:

. . . the Commission found that the company had advertised that these cigarettes and the smoke therefrom contain less nicotine than any of the six other leading brands of cigarettes and that the smoke contains less tars and resins and is less irritating to the throat than cigarettes of the other leading brands, and that the advertising was false, misleading and deceptive. The evidence amply supports this finding.

Laboratory tests introduced in evidence show that the difference in content of nicotine, tars and resins of the different leading brands of cigarettes is insignificant in amount; and there is abundant testimony of medical experts that such difference as there is could result in no difference in the physiological effect upon the smoker. There is expert evidence, also,

that the slight difference in the nicotine, tar and resin content of cigarettes is not constant between different brands, but varies from place to place and from time to time, and that it is a practical impossibility for the manufacturer of cigarettes to determine or to remove or substantially reduce such content or to maintain constancy of such content in the finished cigarette. This testimony gives ample support to the Commission's findings.

The company introduced no evidence in the case but asks that we disregard the testimony of the expert witness who testified to the impossibility of determining, removing or substantially reducing the nicotine, tar or resin content of cigarettes, on the ground that he had had no experience in the manufacturing or blending of tobacco. The record shows, however, that this witness, Dr. McMurtry, is a plant physiologist with the U. S. Department of Agriculture in the Division of Tobacco Investigation and that he has been so employed since 1917. It would seem that his testimony with respect to a matter of this sort should have great weight; but, of course, the weight to be accorded it is a matter for the Commission, not for us, and the Commission believed it. Even if his testimony be disregarded, there remains the testimony of the experts to the effect that the difference in the nicotine, tar and resin content of cigarettes of the leading brands is insignificant and not sufficient to make any difference in the physiological effect upon the smoker. This of itself is sufficient to condemn the advertising as false and misleading, since it is intended to appeal to those who are interested in the physiological effect of the smoke of the cigarettes and who would be led by the advertising to believe that the smoke of the Old Gold cigarettes is less harmful to the smoker because [they] contain appreciably less nicotine, tars and resins.

The company relies upon the truth of the advertisements complained of, saying that they merely state what had been truthfully stated in an article in the *Reader's Digest*. An examination of the advertisements, however, shows a perversion of the meaning of the *Reader's Digest* article which does little credit to the company's advertising department—a perversion which results in the use of the truth in such a way as to cause the reader to believe the exact opposite of what was intended by the writer of the article. A comparison of the advertisements with the article makes this very plain. The article, after referring to laboratory tests that had been made on cigarettes of the leading brands, says:

"The laboratory's general conclusion will be sad news for the advertising copy writers, but good news for the smoker, who need no longer worry as to which cigarette can most effectively nail down his coffin. For one nail is just about as good as another. Says the laboratory report: The differences between brands are, practically speaking, small, and no single brand is so superior to its competitors as to justify its selection on the ground that it is less harmful. How small the variations are may be seen from the data tabulated on page 7."

The table referred to in the article was inserted for the express purpose of showing the insignificance of the difference in the nicotine and tar content of the smoke from the various brands of cigarettes. It appears therefrom that the Old Gold cigarettes examined in the test contained less nicotine, tars and resins than the others examined, although the difference, according to the uncontradicted expert evidence, was so small as to be entirely insignificant and utterly without meaning so far as effect upon the smoker is concerned. The company

proceeded to advertise this difference as though it had received a citation for public service instead of a castigation from the *Reader's Digest*. In the leading newspapers of the country and over the radio it advertised that the *Reader's Digest* had had experiments conducted and had found that Old Gold cigarettes were lowest in nicotine and lowest in irritating tars and resins, just as though a substantial difference in such content had been found. The following advertisement may be taken as typical:

"OLD GOLDS FOUND LOWEST IN NICOTINE; OLD GOLDS FOUND LOWEST IN THROAT-IRRITATING TARS AND RESINS"

"See Impartial Test by *Reader's Digest* July Issue."

"See How Your Brand Compares with Old Gold."

"*Reader's Digest* assigned a scientific testing laboratory to find out about cigarettes. They tested seven leading cigarettes and *Reader's Digest* published the results."

"The cigarette whose smoke was lowest in nicotine was Old Gold. The cigarette with the least throat-irritating tars and resins was Old Gold."

"On both these major counts Old Gold was best among all seven cigarettes tested."

"Get July *Reader's Digest*. Turn to Page 5. See what this highly respected magazine reports."

"You'll say, 'From now on, my cigarette is Old Gold.' Light one? Note the mild, interesting flavor. Easier on the throat? Sure: And more smoking pleasure: Yes, it's the new Old Gold—finer yet, since 'something new has been added'."

The fault with this advertising was not that it did not print all that the *Reader's Digest* article said, but that it printed a small part thereof in such a way as to create an entirely false and misleading impression, not only as to what was said in the article, but also as to the quality of the company's cigarettes. Almost anyone reading the advertisements or listening to the radio broadcasts would have gained the very definite impression that Old Gold cigarettes were less irritating to the throat and less harmful than other leading brands of cigarettes because they contained substantially less nicotine, tars and resins, and that the *Reader's Digest* had established this fact in impartial laboratory tests; and few would have troubled to look up the *Reader's Digest* to see what it really had said. The truth was exactly the opposite. There was no substantial difference in Old Gold cigarettes and the other leading brands with respect to their content of nicotine, tars and resins and this was what the Reader's Digest article plainly said. The table whose meaning the advertisements distorted for the purpose of misleading and deceiving the public was intended to prove that there was no practical difference and did prove it when properly understood. To tell less than the whole truth is a well known method of deception; and he who deceives by resorting to such method cannot excuse the deception by relying upon the truthfulness per se of the partial truth by which it has been accomplished.

In determining whether or not advertising is false or misleading within the meaning of the statute, regard must be had, not to fine spun distinctions and arguments that may be made in excuse, but to the effect which it might reasonably be expected to have upon the general public. "The important criterion is the net impression which the advertisement is likely to make upon the general populace." ... with reference to the law relating to trademarks:

"The law is not made for the protection of experts, but for the public—that vast multitude which includes the ignorant, the unthinking and the credulous, who, in making purchases, do not stop to analyze, but are governed by appearances and general impressions." ... We think that the Commission's determination here was reasonable and amply supported by the evidence before it, and that its order forbidding the advertising as false and misleading was well within the limits of its discretion.

# Chapter 11

# Inferring Causality

# Chapter 12

# Simpson's Paradox

## 12.1 Suggested Reading on Simpson's Paradox

C.4 Simpson's Paradox

C.4.1 – When Combined Data Reveal the Flaw of Averages (Cari Tuna, *Wall Street Journal*, December 2, 2009)

# Chapter 13

# Meta-analysis

## 13.1 Suggested Reading on Meta-analysis

There are two short cautionary pieces on meta-analysis that have appeared over the years in *Science*, and which are listed in the Suggested Reading. One (Section C.5.1) is a news item from Charles Mann, *Meta-Analysis in the Breech* (August 3, 1990). The second in Section C.5.7 is a book review by John C. Bailar III, entitled *Assessing Assessments* (July 25, 1997). The review is for a book by Morton Hunt, *How Science Takes Stock: A History of Meta-Analysis* (Russell Sage Foundation, 1997).

C.5 – Meta-Analysis

C.5.1 – Meta-Analysis in the Breech (Charles Mann, *Science*, August 3, 1990)

C.5.2 – The Promise and Problems of Meta-Analysis (John C. Bailar III, *New England Journal of Medicine*, August 21, 1997)

C.5.3 – Meta-Analyses and Large Randomized, Controlled Trials (*New England Journal of Medicine*, various letters to the Editor, and rejoinders, 1997)

C.5.4 – Primary, Secondary, and Meta-Analysis of Research (Gene Glass, *Educational Researcher*, *5(10)*, 3–8)

C.5.5 – Meta-Analysis at 25 (Gene Glass, January 2000; available online)

C.5.6 – Meta-Analysis and Depression: A Recent Cautionary Example

Popular Drugs May Help Only Severe Depression (Benedict Carey, *New York Times*, January 6, 2010)

The Wrong Story About Depression (Judith Warner, *New York Times*, January 9, 2010)

Before You Quit Antidepressants ... (Richard A. Friedman, *New York Times*, January 12, 2010)

C.5.7 – Assessing Assessments (John C. Bailar III, *Science*, July 25, 1997)

# Chapter 14

# Statistical Sleuthing and Explanation

## 14.1 Suggested Reading on Statistical Sleuthing

C.6 – Sleuthing

C.6.1 – Following Benford's Law, or Looking Out for No. 1 (Malcom W. Browne, *New York Times*, August 4, 1990)

C.6.2 – Rise and Flaw of Internet's Election-Fraud Hunters (Carl Bialik, *Wall Street Journal*, July 1, 2009)

C.6.3 – A Little Ignorance: How Statistics Rescued a Damsel in Distress (Peter Baldwin and Howard Wainer, *Chance*, 2009, 49–52)

C.6.4 – The Ghost's Vocabulary: How the Computer Listens for Shakespeare's "Voiceprint" (Edward Dolnick, *The Atlantic*, October, 1991)

C.6.5 – The Environment and Disease: Association or Causation? (Austin Bradford Hill, *Proceedings of the Royal Society of Medicine*, *58*, 1965, 295-300)

### 14.1.1 Appendix: U.S. Supreme Court, McCleskey v. Kemp (Decided: April 22, 1987): Majority Opinion and Dissent

Justice Powell delivered the opinion of the Court.

This case presents the question whether a complex statistical study that indicates a risk that racial considerations enter into capital sentencing determinations proves that petitioner McCleskey's capital sentence is unconstitutional under the Eighth or Fourteenth Amendment.

. . . McCleskey next filed a petition for a writ of habeas corpus in the Federal District Court for the Northern District of Georgia. His petition raised 18 claims, one of which was that the Georgia capital sentencing process is administered in a racially discriminatory manner in violation of the Eighth and Fourteenth Amendments to the United States Constitution. In support of his claim, McCleskey proffered a statistical study performed by Professors David C. Baldus, Charles Pulaski, and George Woodworth (the Baldus study) that purports to show a disparity in the imposition of the death sentence in Georgia based on

the race of the murder victim and, to a lesser extent, the race of the defendant. The Baldus study is actually two sophisticated statistical studies that examine over 2,000 murder cases that occurred in Georgia during the 1970's. The raw numbers collected by Professor Baldus indicate that defendants charged with killing white persons received the death penalty in 11% of the cases, but defendants charged with killing blacks received the death penalty in only 1% of the cases. The raw numbers also indicate a reverse racial disparity according to the race of the defendant: 4% of the black defendants received the death penalty, as opposed to 7% of the white defendants.

Baldus also divided the cases according to the combination of the race of the defendant and the race of the victim. He found that the death penalty was assessed in 22% of the cases involving black defendants and white victims; 8% of the cases involving white defendants and white victims; 1% of the cases involving black defendants and black victims; and 3% of the cases involving white defendants and black victims. Similarly, Baldus found that prosecutors sought the death penalty in 70% of the cases involving black defendants and white victims; 32% of the cases involving white defendants and white victims; 15% of the cases involving black defendants and black victims; and 19% of the cases involving white defendants and black victims.

Baldus subjected his data to an extensive analysis, taking account of 230 variables that could have explained the disparities on nonracial grounds. One of his models concludes that, even after taking account of 39 nonracial variables, defendants charged with killing white victims were 4.3 times as likely to receive a death sentence as defendants charged with killing blacks. According to this model, black defendants were 1.1 times as likely to receive a death sentence as other defendants. Thus, the Baldus study indicates that black defendants, such as McCleskey, who kill white victims have the greatest likelihood of receiving the death penalty.

The District Court held an extensive evidentiary hearing on McCleskey's petition. ... It concluded that McCleskey's statistics do not demonstrate a prima facie case in support of the contention that the death penalty was imposed upon him because of his race, because of the race of the victim, or because of any Eighth Amendment concern.

As to McCleskey's Fourteenth Amendment claim, the court found that the methodology of the Baldus study was flawed in several respects. Because of these defects, the court held that the Baldus study "fail[ed] to contribute anything of value" to McCleskey's claim. Accordingly, the court denied the petition insofar as it was based upon the Baldus study.[1]

[1]Baldus, among other experts, testified at the evidentiary hearing. The District Court "was impressed with the learning of all of the experts." Nevertheless, the District Court noted that, in many respects, the data were incomplete. In its view, the questionnaires used to obtain the data failed to capture the full degree of the aggravating or mitigating circumstances. The court criticized the researcher's decisions regarding unknown variables. The researchers could not discover whether penalty trials were held in many of the cases, thus undercutting the value of the study's statistics as to prosecutorial decisions. In certain cases, the study lacked information on the race of the victim in cases involving multiple victims, on whether or not the prosecutor offered a plea bargain, and on credibility problems with witnesses. The court concluded that McCleskey had failed to establish by a preponderance of the evidence that the data were trustworthy.

It is a major premise of a statistical case that the database numerically mirrors reality. If it does not in

The Court of Appeals for the Eleventh Circuit, sitting en banc, carefully reviewed the District Court's decision on McCleskey's claim. It assumed the validity of the study itself, and addressed the merits of McCleskey's Eighth and Fourteenth Amendment claims. That is, the court assumed that the study showed that systematic and substantial disparities existed in the penalties imposed upon homicide defendants in Georgia based on race of the homicide victim, that the disparities existed at a less substantial rate in death sentencing based on race of defendants, and that the factors of race of the victim and defendant were at work in Fulton County.

Even assuming the study's validity, the Court of Appeals found the statistics insufficient to demonstrate discriminatory intent or unconstitutional discrimination in the Fourteenth Amendment context, [and] insufficient to show irrationality, arbitrariness and capriciousness under any kind of Eighth Amendment analysis.

The court noted:

The very exercise of discretion means that persons exercising discretion may reach different results from exact duplicates. Assuming each result is within the range of discretion, all are correct in the eyes of the law. It would not make sense for the system to require the exercise of discretion in order to be facially constitutional, and at the same time hold a system unconstitutional in application where that discretion achieved different results for what appear to be exact duplicates, absent the state showing the reasons for the difference.

The Baldus approach ... would take the cases with different results on what are contended to be duplicate facts, where the differences could not be otherwise explained, and conclude that the different result was based on race alone. ... This approach ignores the realities. ... There are, in fact, no exact duplicates in capital crimes and capital defendants. The type of research submitted here tends to show which of the directed factors were effective, but is of restricted use in showing what undirected factors control the exercise of constitutionally required discretion.

The court concluded:

Viewed broadly, it would seem that the statistical evidence presented here, assuming its

---

substantial degree mirror reality, any inferences empirically arrived at are untrustworthy.

The District Court noted other problems with Baldus' methodology. First, the researchers assumed that all of the information available from the questionnaires was available to the juries and prosecutors when the case was tried. The court found this assumption "questionable." Second, the court noted the instability of the various models. Even with the 230-variable model, consideration of 20 further variables caused a significant drop in the statistical significance of race. In the court's view, this undermined the persuasiveness of the model that showed the greatest racial disparity, the 39-variable model. Third, the court found that the high correlation between race and many of the nonracial variables diminished the weight to which the study was entitled.

Finally, the District Court noted the inability of any of the models to predict the outcome of actual cases. As the court explained, statisticians use a measure called an "r-squared" to measure what portion of the variance in the dependent variable (death sentencing rate, in this case) is accounted for by the independent variables of the model. A perfectly predictive model would have an r-squared value of 1.0. A model with no predictive power would have an r-squared value of 0. The r-squared value of Baldus' most complex model, the 230-variable model, was between .46 and .48. Thus, as the court explained, "the 230-variable model does not predict the outcome in half of the cases."

validity, confirms, rather than condemns, the system. ... The marginal disparity based on the race of the victim tends to support the state's contention that the system is working far differently from the one which Furman v. Georgia, condemned. In pre-Furman days, there was no rhyme or reason as to who got the death penalty and who did not. But now, in the vast majority of cases, the reasons for a difference are well documented. That they are not so clear in a small percentage of the cases is no reason to declare the entire system unconstitutional.

The Court of Appeals affirmed the denial by the District Court of McCleskey's petition for a writ of habeas corpus insofar as the petition was based upon the Baldus study, with three judges dissenting as to McCleskey's claims based on the Baldus study. We granted certiorari, and now affirm.

... McCleskey's first claim is that the Georgia capital punishment statute violates the Equal Protection Clause of the Fourteenth Amendment.[2] He argues that race has infected the administration of Georgia's statute in two ways: persons who murder whites are more likely to be sentenced to death than persons who murder blacks, and black murderers are more likely to be sentenced to death than white murderers. As a black defendant who killed a white victim, McCleskey claims that the Baldus study demonstrates that he was discriminated against because of his race and because of the race of his victim. In its broadest form, McCleskey's claim of discrimination extends to every actor in the Georgia capital sentencing process, from the prosecutor who sought the death penalty and the jury that imposed the sentence to the State itself that enacted the capital punishment statute and allows it to remain in effect despite its allegedly discriminatory application. We agree with the Court of Appeals, and every other court that has considered such a challenge, that this claim must fail.

Our analysis begins with the basic principle that a defendant who alleges an equal protection violation has the burden of proving "the existence of purposeful discrimination." A corollary to this principle is that a criminal defendant must prove that the purposeful discrimination "had a discriminatory effect" on him. Thus, to prevail under the Equal Protection Clause, McCleskey must prove that the decision-makers in his case acted with discriminatory purpose. He offers no evidence specific to his own case that would support an inference that racial considerations played a part in his sentence. Instead, he relies solely on the Baldus study. McCleskey argues that the Baldus study compels an inference that his sentence rests on purposeful discrimination. McCleskey's claim that these statistics are sufficient proof of discrimination, without regard to the facts of a particular case, would extend to all capital cases in Georgia, at least where the victim was white and the defendant is black.

[2]Although the District Court rejected the findings of the Baldus study as flawed, the Court of Appeals assumed that the study is valid, and reached the constitutional issues. Accordingly, those issues are before us. As did the Court of Appeals, we assume the study is valid statistically, without reviewing the factual findings of the District Court. Our assumption that the Baldus study is statistically valid does not include the assumption that the study shows that racial considerations actually enter into any sentencing decisions in Georgia. Even a sophisticated multiple-regression analysis such as the Baldus study can only demonstrate a risk that the factor of race entered into some capital sentencing decisions, and a necessarily lesser risk that race entered into any particular sentencing decision.

The Court has accepted statistics as proof of intent to discriminate in certain limited contexts. First, this Court has accepted statistical disparities as proof of an equal protection violation in the selection of the jury venire in a particular district. Although statistical proof normally must present a "stark" pattern to be accepted as the sole proof of discriminatory intent under the Constitution, [b]ecause of the nature of the jury-selection task, ... we have permitted a finding of constitutional violation even when the statistical pattern does not approach [such] extremes.

Second, this Court has accepted statistics in the form of multiple-regression analysis to prove statutory violations under Title VII of the Civil Rights Act of 1964.

But the nature of the capital sentencing decision, and the relationship of the statistics to that decision, are fundamentally different from the corresponding elements in the venire selection or Title VII cases. Most importantly, each particular decision to impose the death penalty is made by a petit jury selected from a properly constituted venire. Each jury is unique in its composition, and the Constitution requires that its decision rest on consideration of innumerable factors that vary according to the characteristics of the individual defendant and the facts of the particular capital offense. Thus, the application of an inference drawn from the general statistics to a specific decision in a trial and sentencing simply is not comparable to the application of an inference drawn from general statistics to a specific venire-selection or Title VII case. In those cases, the statistics relate to fewer entities, and fewer variables are relevant to the challenged decisions.

Another important difference between the cases in which we have accepted statistics as proof of discriminatory intent and this case is that, in the venire-selection and Title VII contexts, the decision-maker has an opportunity to explain the statistical disparity. Here, the State has no practical opportunity to rebut the Baldus study. "[C]ontrolling considerations of ... public policy," dictate that jurors "cannot be called ... to testify to the motives and influences that led to their verdict." Similarly, the policy considerations behind a prosecutor's traditionally "wide discretion" suggest the impropriety of our requiring prosecutors to defend their decisions to seek death penalties, "often years after they were made." Moreover, absent far stronger proof, it is unnecessary to seek such a rebuttal, because a legitimate and unchallenged explanation for the decision is apparent from the record: McCleskey committed an act for which the United States Constitution and Georgia laws permit imposition of the death penalty.

Finally, McCleskey's statistical proffer must be viewed in the context of his challenge. McCleskey challenges decisions at the heart of the State's criminal justice system.

[O]ne of society's most basic tasks is that of protecting the lives of its citizens, and one of the most basic ways in which it achieves the task is through criminal laws against murder.

Implementation of these laws necessarily requires discretionary judgments. Because discretion is essential to the criminal justice process, we would demand exceptionally clear proof before we would infer that the discretion has been abused. The unique nature of the decisions at issue in this case also counsels against adopting such an inference from the disparities indicated by the Baldus study. Accordingly, we hold that the Baldus study is clearly insufficient to support an inference that any of the decision-makers in McCleskey's

case acted with discriminatory purpose.

. . . McCleskey also suggests that the Baldus study proves that the State as a whole has acted with a discriminatory purpose. He appears to argue that the State has violated the Equal Protection Clause by adopting the capital punishment statute and allowing it to remain in force despite its allegedly discriminatory application. But "[d]iscriminatory purpose" . . . implies more than intent as volition or intent as awareness of consequences. It implies that the decision-maker, in this case a state legislature, selected or reaffirmed a particular course of action at least in part "because of," not merely "in spite of," its adverse effects upon an identifiable group.

For this claim to prevail, McCleskey would have to prove that the Georgia Legislature enacted or maintained the death penalty statute because of an anticipated racially discriminatory effect. In Gregg v. Georgia, this Court found that the Georgia capital sentencing system could operate in a fair and neutral manner. There was no evidence then, and there is none now, that the Georgia Legislature enacted the capital punishment statute to further a racially discriminatory purpose. Nor has McCleskey demonstrated that the legislature maintains the capital punishment statute because of the racially disproportionate impact suggested by the Baldus study. As legislatures necessarily have wide discretion in the choice of criminal laws and penalties, and as there were legitimate reasons for the Georgia Legislature to adopt and maintain capital punishment, we will not infer a discriminatory purpose on the part of the State of Georgia. Accordingly, we reject McCleskey's equal protection claims.

. . .

Although our decision in Gregg as to the facial validity of the Georgia capital punishment statute appears to foreclose McCleskey's disproportionality argument, he further contends that the Georgia capital punishment system is arbitrary and capricious in application, and therefore his sentence is excessive, because racial considerations may influence capital sentencing decisions in Georgia. We now address this claim.

To evaluate McCleskey's challenge, we must examine exactly what the Baldus study may show. Even Professor Baldus does not contend that his statistics prove that race enters into any capital sentencing decisions, or that race was a factor in McCleskey's particular case. Statistics, at most, may show only a likelihood that a particular factor entered into some decisions. There is, of course, some risk of racial prejudice influencing a jury's decision in a criminal case. There are similar risks that other kinds of prejudice will influence other criminal trials. The question "is at what point that risk becomes constitutionally unacceptable," McCleskey asks us to accept the likelihood allegedly shown by the Baldus study as the constitutional measure of an unacceptable risk of racial prejudice influencing capital sentencing decisions. This we decline to do. Because of the risk that the factor of race may enter the criminal justice process, we have engaged in "unceasing efforts" to eradicate racial prejudice from our criminal justice system. Our efforts have been guided by our recognition that the inestimable privilege of trial by jury . . . is a vital principle, underlying the whole administration of criminal justice system. Thus, it is the jury that is a criminal defendant's fundamental "protection of life and liberty against race or color prejudice." Specifically, a

capital sentencing jury representative of a criminal defendant's community assures a 'diffused impartiality,' in the jury's task of "express[ing] the conscience of the community on the ultimate question of life or death.

Individual jurors bring to their deliberations "qualities of human nature and varieties of human experience, the range of which is unknown and perhaps unknowable." The capital sentencing decision requires the individual jurors to focus their collective judgment on the unique characteristics of a particular criminal defendant. It is not surprising that such collective judgments often are difficult to explain. But the inherent lack of predictability of jury decisions does not justify their condemnation. On the contrary, it is the jury's function to make the difficult and uniquely human judgments that defy codification, and that "buil[d] discretion, equity, and flexibility into a legal system."

McCleskey's argument that the Constitution condemns the discretion allowed decisionmakers in the Georgia capital sentencing system is antithetical to the fundamental role of discretion in our criminal justice system. Discretion in the criminal justice system offers substantial benefits to the criminal defendant. Not only can a jury decline to impose the death sentence, it can decline to convict or choose to convict of a lesser offense. Whereas decisions against a defendant's interest may be reversed by the trial judge or on appeal, these discretionary exercises of leniency are final and unreviewable. Similarly, the capacity of prosecutorial discretion to provide individualized justice is "only entrenched in American law." As we have noted, a prosecutor can decline to charge, offer a plea bargain, or decline to seek a death sentence in any particular case. Of course, "the power to be lenient [also] is the power to discriminate," but a capital punishment system that did not allow for discretionary acts of leniency "would be totally alien to our notions of criminal justice."

. . . At most, the Baldus study indicates a discrepancy that appears to correlate with race. Apparent disparities in sentencing are an inevitable part of our criminal justice system. The discrepancy indicated by the Baldus study is "a far cry from the major systemic defects identified in Furman." As this Court has recognized, any mode for determining guilt or punishment "has its weaknesses and the potential for misuse." Specifically, "there can be 'no perfect procedure for deciding in which cases governmental authority should be used to impose death.'" Despite these imperfections, our consistent rule has been that constitutional guarantees are met when "the mode [for determining guilt or punishment] itself has been surrounded with safeguards to make it as fair as possible." Where the discretion that is fundamental to our criminal process is involved, we decline to assume that what is unexplained is invidious. In light of the safeguards designed to minimize racial bias in the process, the fundamental value of jury trial in our criminal justice system, and the benefits that discretion provides to criminal defendants, we hold that the Baldus study does not demonstrate a constitutionally significant risk of racial bias affecting the Georgia capital sentencing process.

. . . Two additional concerns inform our decision in this case. First, McCleskey's claim, taken to its logical conclusion, throws into serious question the principles that underlie our entire criminal justice system. The Eighth Amendment is not limited in application to capital punishment, but applies to all penalties. Thus, if we accepted McCleskey's claim

that racial bias has impermissibly tainted the capital sentencing decision, we could soon be faced with similar claims as to other types of penalty. Moreover, the claim that his sentence rests on the irrelevant factor of race easily could be extended to apply to claims based on unexplained discrepancies that correlate to membership in other minority groups, and even to gender. Similarly, since McCleskey's claim relates to the race of his victim, other claims could apply with equally logical force to statistical disparities that correlate with the race or sex of other actors in the criminal justice system, such as defense attorneys or judges. Also, there is no logical reason that such a claim need be limited to racial or sexual bias. If arbitrary and capricious punishment is the touchstone under the Eighth Amendment, such a claim could—at least in theory—be based upon any arbitrary variable, such as the defendant's facial characteristics, or the physical attractiveness of the defendant or the victim, that some statistical study indicates may be influential in jury decision-making. As these examples illustrate, there is no limiting principle to the type of challenge brought by McCleskey. The Constitution does not require that a State eliminate any demonstrable disparity that correlates with a potentially irrelevant factor in order to operate a criminal justice system that includes capital punishment. As we have stated specifically in the context of capital punishment, the Constitution does not "plac[e] totally unrealistic conditions on its use."

Second, McCleskey's arguments are best presented to the legislative bodies. It is not the responsibility—or indeed even the right—of this Court to determine the appropriate punishment for particular crimes. It is the legislatures, the elected representatives of the people, that are "constituted to respond to the will and consequently the moral values of the people." Legislatures also are better qualified to weigh and evaluate the results of statistical studies in terms of their own local conditions and with a flexibility of approach that is not available to the courts,

Capital punishment is now the law in more than two-thirds of our States. It is the ultimate duty of courts to determine on a case-by-case basis whether these laws are applied consistently with the Constitution. Despite McCleskey's wide-ranging arguments that basically challenge the validity of capital punishment in our multiracial society, the only question before us is whether, in his case, the law of Georgia was properly applied. We agree with the District Court and the Court of Appeals for the Eleventh Circuit that this was carefully and correctly done in this case.

Accordingly, we affirm the judgment of the Court of Appeals for the Eleventh Circuit.

It is so ordered.

---

Justice Brennan, Dissenting Opinion

. . .

At some point in this case, Warren McCleskey doubtless asked his lawyer whether a jury was likely to sentence him to die. A candid reply to this question would have been disturbing. First, counsel would have to tell McCleskey that few of the details of the crime or of McCleskey's past criminal conduct were more important than the fact that his victim was white. Furthermore, counsel would feel bound to tell McCleskey that defendants charged

with killing white victims in Georgia are 4.3 times as likely to be sentenced to death as defendants charged with killing blacks. In addition, frankness would compel the disclosure that it was more likely than not that the race of McCleskey's victim would determine whether he received a death sentence: 6 of every 11 defendants convicted of killing a white person would not have received the death penalty if their victims had been black, while, among defendants with aggravating and mitigating factors comparable to McCleskey's, 20 of every 34 would not have been sentenced to die if their victims had been black. Finally, the assessment would not be complete without the information that cases involving black defendants and white victims are more likely to result in a death sentence than cases featuring any other racial combination of defendant and victim. The story could be told in a variety of ways, but McCleskey could not fail to grasp its essential narrative line: there was a significant chance that race would play a prominent role in determining if he lived or died.

The Court today holds that Warren McCleskey's sentence was constitutionally imposed. It finds no fault in a system in which lawyers must tell their clients that race casts a large shadow on the capital sentencing process. The Court arrives at this conclusion by stating that the Baldus study cannot "prove that race enters into any capital sentencing decisions or that race was a factor in McCleskey's particular case." Since, according to Professor Baldus, we cannot say "to a moral certainty" that race influenced a decision, we can identify only "a likelihood that a particular factor entered into some decisions," and "a discrepancy that appears to correlate with race." This "likelihood" and "discrepancy," holds the Court, is insufficient to establish a constitutional violation. The Court reaches this conclusion by placing four factors on the scales opposite McCleskey's evidence: the desire to encourage sentencing discretion, the existence of "statutory safeguards" in the Georgia scheme, the fear of encouraging widespread challenges to other sentencing decisions, and the limits of the judicial role. The Court's evaluation of the significance of petitioner's evidence is fundamentally at odds with our consistent concern for rationality in capital sentencing, and the considerations that the majority invokes to discount that evidence cannot justify ignoring its force.

. . . It is important to emphasize at the outset that the Court's observation that McCleskey cannot prove the influence of race on any particular sentencing decision is irrelevant in evaluating his Eighth Amendment claim. Since Furman v. Georgia, the Court has been concerned with the risk of the imposition of an arbitrary sentence, rather than the proven fact of one. Furman held that the death penalty may not be imposed under sentencing procedures that create a substantial risk that the punishment will be inflicted in an arbitrary and capricious manner.

As Justice O'Connor observed in Caldwell v. Mississippi, a death sentence must be struck down when the circumstances under which it has been imposed creat[e] an unacceptable risk that "the death penalty [may have been] meted out arbitrarily or capriciously," or through "whim or mistake." This emphasis on risk acknowledges the difficulty of divining the jury's motivation in an individual case. In addition, it reflects the fact that concern for arbitrariness focuses on the rationality of the system as a whole, and that a system that features a significant probability that sentencing decisions are influenced by impermissible

considerations cannot be regarded as rational. As we said in Gregg v. Georgia, "the petitioner looks to the sentencing system as a whole (as the Court did in Furman and we do today)": a constitutional violation is established if a plaintiff demonstrates a "pattern of arbitrary and capricious sentencing."

As a result, our inquiry under the Eighth Amendment has not been directed to the validity of the individual sentences before us. In Godfrey, for instance, the Court struck down the petitioner's sentence because the vagueness of the statutory definition of heinous crimes created a risk that prejudice or other impermissible influences might have infected the sentencing decision. In vacating the sentence, we did not ask whether it was likely that Godfrey's own sentence reflected the operation of irrational considerations. Nor did we demand a demonstration that such considerations had actually entered into other sentencing decisions involving heinous crimes. Similarly, in Roberts v. Louisiana, and Woodson v. North Carolina, we struck down death sentences in part because mandatory imposition of the death penalty created the risk that a jury might rely on arbitrary considerations in deciding which persons should be convicted of capital crimes. Such a risk would arise, we said, because of the likelihood that jurors, reluctant to impose capital punishment on a particular defendant, would refuse to return a conviction, so that the effect of mandatory sentencing would be to recreate the unbounded sentencing discretion condemned in Furman. We did not ask whether the death sentences in the cases before us could have reflected the jury's rational consideration and rejection of mitigating factors. Nor did we require proof that juries had actually acted irrationally in other cases.

Defendants challenging their death sentences thus never have had to prove that impermissible considerations have actually infected sentencing decisions. We have required instead that they establish that the system under which they were sentenced posed a significant risk of such an occurrence. McCleskey's claim does differ, however, in one respect from these earlier cases: it is the first to base a challenge not on speculation about how a system might operate, but on empirical documentation of how it does operate.

The Court assumes the statistical validity of the Baldus study, and acknowledges that McCleskey has demonstrated a risk that racial prejudice plays a role in capital sentencing in Georgia. Nonetheless, it finds the probability of prejudice insufficient to create constitutional concern. Close analysis of the Baldus study, however, in light of both statistical principles and human experience, reveals that the risk that race influenced McCleskey's sentence is intolerable by any imaginable standard.

. . . The Baldus study indicates that, after taking into account some 230 nonracial factors that might legitimately influence a sentencer, the jury more likely than not would have spared McCleskey's life had his victim been black. The study distinguishes between those cases in which (1) the jury exercises virtually no discretion because the strength or weakness of aggravating factors usually suggests that only one outcome is appropriate; and (2) cases reflecting an "intermediate" level of aggravation, in which the jury has considerable discretion in choosing a sentence. McCleskey's case falls into the intermediate range. In such cases, death is imposed in 34% of white-victim crimes and 14% of black-victim crimes, a difference of 139% in the rate of imposition of the death penalty. In other words, just under 59%—

almost 6 in 10—defendants comparable to McCleskey would not have received the death penalty if their victims had been black.

Furthermore, even examination of the sentencing system as a whole, factoring in those cases in which the jury exercises little discretion, indicates the influence of race on capital sentencing. For the Georgia system as a whole, race accounts for a six percentage point difference in the rate at which capital punishment is imposed. Since death is imposed in 11% of all white-victim cases, the rate in comparably aggravated black-victim cases is 5%. The rate of capital sentencing in a white-victim case is thus 120% greater than the rate in a black-victim case. Put another way, over half—55%—of defendants in white-victim crimes in Georgia would not have been sentenced to die if their victims had been black. Of the more than 200 variables potentially relevant to a sentencing decision, race of the victim is a powerful explanation for variation in death sentence rates—as powerful as nonracial aggravating factors such as a prior murder conviction or acting as the principal planner of the homicide.

These adjusted figures are only the most conservative indication of the risk that race will influence the death sentences of defendants in Georgia. Data unadjusted for the mitigating or aggravating effect of other factors show an even more pronounced disparity by race. The capital sentencing rate for all white-victim cases was almost 11 times greater than the rate for black-victim cases. Furthermore, blacks who kill whites are sentenced to death at nearly 22 times the rate of blacks who kill blacks, and more than 7 times the rate of whites who kill blacks. In addition, prosecutors seek the death penalty for 70% of black defendants with white victims, but for only 15% of black defendants with black victims, and only 19% of white defendants with black victims. Since our decision upholding the Georgia capital sentencing system in Gregg, the State has executed seven persons. All of the seven were convicted of killing whites, and six of the seven executed were black. Such execution figures are especially striking in light of the fact that, during the period encompassed by the Baldus study, only 9.2% of Georgia homicides involved black defendants and white victims, while 60.7% involved black victims.

McCleskey's statistics have particular force because most of them are the product of sophisticated multiple-regression analysis. Such analysis is designed precisely to identify patterns in the aggregate, even though we may not be able to reconstitute with certainty any individual decision that goes to make up that pattern. Multiple-regression analysis is particularly well suited to identify the influence of impermissible considerations in sentencing, since it is able to control for permissible factors that may explain an apparent arbitrary pattern. While the decision-making process of a body such as a jury may be complex, the Baldus study provides a massive compilation of the details that are most relevant to that decision. As we held in the context of Title VII of the Civil Rights Act of 1964 last Term in Bazemore v. Friday, a multiple-regression analysis need not include every conceivable variable to establish a party's case, as long as it includes those variables that account for the major factors that are likely to influence decisions. In this case, Professor Baldus in fact conducted additional regression analyses in response to criticisms and suggestions by the District Court, all of which confirmed, and some of which even strengthened, the study's

original conclusions.

The statistical evidence in this case thus relentlessly documents the risk that McCleskey's sentence was influenced by racial considerations. This evidence shows that there is a better than even chance in Georgia that race will influence the decision to impose the death penalty: a majority of defendants in white-victim crimes would not have been sentenced to die if their victims had been black. In determining whether this risk is acceptable, our judgment must be shaped by the awareness that [t]he risk of racial prejudice infecting a capital sentencing proceeding is especially serious in light of the complete finality of the death sentence, and that [i]t is of vital importance to the defendant and to the community that any decision to impose the death sentence be, and appear to be, based on reason rather than caprice or emotion. In determining the guilt of a defendant, a State must prove its case beyond a reasonable doubt. That is, we refuse to convict if the chance of error is simply less likely than not. Surely, we should not be willing to take a person's life if the chance that his death sentence was irrationally imposed is more likely than not. In light of the gravity of the interest at stake, petitioner's statistics, on their face, are a powerful demonstration of the type of risk that our Eighth Amendment jurisprudence has consistently condemned.

. . . Evaluation of McCleskey's evidence cannot rest solely on the numbers themselves. We must also ask whether the conclusion suggested by those numbers is consonant with our understanding of history and human experience. Georgia's legacy of a race-conscious criminal justice system, as well as this Court's own recognition of the persistent danger that racial attitudes may affect criminal proceedings, indicates that McCleskey's claim is not a fanciful product of mere statistical artifice.

For many years, Georgia operated openly and formally precisely the type of dual system the evidence shows is still effectively in place. The criminal law expressly differentiated between crimes committed by and against blacks and whites, distinctions whose lineage traced back to the time of slavery. During the colonial period, black slaves who killed whites in Georgia, regardless of whether in self-defense or in defense of another, were automatically executed.

By the time of the Civil War, a dual system of crime and punishment was well established in Georgia. The state criminal code contained separate sections for "Slaves and Free Persons of Color," and for all other persons. The code provided, for instance, for an automatic death sentence for murder committed by blacks, but declared that anyone else convicted of murder might receive life imprisonment if the conviction were founded solely on circumstantial testimony or simply if the jury so recommended. The code established that the rape of a free white female by a black "shall be punishable by death. However, rape by anyone else of a free white female was punishable by a prison term not less than 2 nor more than 20 years. The rape of blacks was punishable "by fine and imprisonment, at the discretion of the court." A black convicted of assaulting a free white person with intent to murder could be put to death at the discretion of the court, but the same offense committed against a black, slave or free, was classified as a "minor" offense whose punishment lay in the discretion of the court, as long as such punishment did not "extend to life, limb, or health." Assault with intent to murder by a white person was punishable by a prison term of from 2 to 10 years.

While sufficient provocation could reduce a charge of murder to manslaughter, the code provided that [o]bedience and submission being the duty of a slave, much greater provocation is necessary to reduce a homicide of a white person by him to voluntary manslaughter, than is prescribed for white persons.

In more recent times, some 40 years ago, Gunnar Myrdal's epochal study of American race relations produced findings mirroring McCleskey's evidence:

As long as only Negroes are concerned and no whites are disturbed, great leniency will be shown in most cases. ... The sentences for even major crimes are ordinarily reduced when the victim is another Negro.

... For offenses which involve any actual or potential danger to whites, however, Negroes are punished more severely than whites.

... On the other hand, it is quite common for a white criminal to be set free if his crime was against a Negro.

... This Court has invalidated portions of the Georgia capital sentencing system three times over the past 15 years. The specter of race discrimination was acknowledged by the Court in striking down the Georgia death penalty statute in Furman. Justice Douglas cited studies suggesting imposition of the death penalty in racially discriminatory fashion, and found the standard-less statutes before the Court "pregnant with discrimination." Justice Marshall pointed to statistics indicating that Negroes [have been] executed far more often than whites in proportion to their percentage of the population. Studies indicate that, while the higher rate of execution among Negroes is partially due to a higher rate of crime, there is evidence of racial discrimination. Although Justice Stewart declined to conclude that racial discrimination had been plainly proved, he stated that [m]y concurring Brothers have demonstrated that, if any basis can be discerned for the selection of these few to be sentenced to die, it is the constitutionally impermissible basis of race. In dissent, Chief Justice Burger acknowledged that statistics suggest, at least as a historical matter, that Negroes have been sentenced to death with greater frequency than whites in several States, particularly for the crime of interracial rape. Finally, also in dissent, Justice Powell intimated that an Equal Protection Clause argument would be available for a black who could demonstrate that members of his race were being singled out for more severe punishment than others charged with the same offense. He noted that, although the Eighth Circuit had rejected a claim of discrimination in Maxwell v. Bishop, vacated and remanded on other grounds, the statistical evidence in that case tend[ed] to show a pronounced disproportion in the number of Negroes receiving death sentences for rape in parts of Arkansas and elsewhere in the South. It is clear that the Court regarded the opportunity for the operation of racial prejudice a particularly troublesome aspect of the unbounded discretion afforded by the Georgia sentencing scheme. Five years later, the Court struck down the imposition of the death penalty in Georgia for the crime of rape. Although the Court did not explicitly mention race, the decision had to have been informed by the specific observations on rape by both the Chief Justice and Justice Powell in Furman. Furthermore, evidence submitted to the Court indicated that black men who committed rape, particularly of white women, were considerably more likely to be sentenced to death than white rapists. For instance, by 1977, Georgia had executed

62 men for rape since the Federal Government began compiling statistics in 1930. Of these men, 58 were black and 4 were white. Three years later, the Court in Godfrey found one of the State's statutory aggravating factors unconstitutionally vague, since it resulted in "standard-less and unchanneled imposition of death sentences in the uncontrolled discretion of a basically uninstructed jury. . . . " Justice Marshall, concurring in the judgment, noted that [t]he disgraceful distorting effects of racial discrimination and poverty continue to be painfully visible in the imposition of death sentences.

This historical review of Georgia criminal law is not intended as a bill of indictment calling the State to account for past transgressions. Citation of past practices does not justify the automatic condemnation of current ones. But it would be unrealistic to ignore the influence of history in assessing the plausible implications of McCleskey's evidence. [A]mericans share a historical experience that has resulted in individuals within the culture ubiquitously attaching a significance to race that is irrational and often outside their awareness. As we said in Rose v. Mitchell: [W]e . . . cannot deny that, 114 years after the close of the War Between the States and nearly 100 years after Strauder, racial and other forms of discrimination still remain a fact of life, in the administration of justice as in our society as a whole. Perhaps today that discrimination takes a form more subtle than before. But it is not less real or pernicious.

The ongoing influence of history is acknowledged, as the majority observes, by our "unceasing efforts to eradicate racial prejudice from our criminal justice system." These efforts, however, signify not the elimination of the problem, but its persistence. Our cases reflect a realization of the myriad of opportunities for racial considerations to influence criminal proceedings: in the exercise of peremptory challenges, in the selection of the grand jury, in the selection of the petit jury, in the exercise of prosecutorial discretion, in the conduct of argument, and in the conscious or unconscious bias of jurors.

The discretion afforded prosecutors and jurors in the Georgia capital sentencing system creates such opportunities. No guidelines govern prosecutorial decisions to seek the death penalty, and Georgia provides juries with no list of aggravating and mitigating factors, nor any standard for balancing them against one another. Once a jury identifies one aggravating factor, it has complete discretion in choosing life or death, and need not articulate its basis for selecting life imprisonment. The Georgia sentencing system therefore provides considerable opportunity for racial considerations, however subtle and unconscious, to influence charging and sentencing decisions.

History and its continuing legacy thus buttress the probative force of McCleskey's statistics. Formal dual criminal laws may no longer be in effect, and intentional discrimination may no longer be prominent. Nonetheless, as we acknowledged in Turner, "subtle, less consciously held racial attitudes" continue to be of concern, and the Georgia system gives such attitudes considerable room to operate. The conclusions drawn from McCleskey's statistical evidence are therefore consistent with the lessons of social experience.

The majority thus misreads our Eighth Amendment jurisprudence in concluding that McCleskey has not demonstrated a degree of risk sufficient to raise constitutional concern. The determination of the significance of his evidence is at its core an exercise in human moral

judgment, not a mechanical statistical analysis. It must first and foremost be informed by awareness of the fact that death is irrevocable, and that, as a result, the qualitative difference of death from all other punishments requires a greater degree of scrutiny of the capital sentencing determination. For this reason, we have demanded a uniquely high degree of rationality in imposing the death penalty. A capital sentencing system in which race more likely than not plays a role does not meet this standard. It is true that every nuance of decision cannot be statistically captured, nor can any individual judgment be plumbed with absolute certainty. Yet the fact that we must always act without the illumination of complete knowledge cannot induce paralysis when we confront what is literally an issue of life and death. Sentencing data, history, and experience all counsel that Georgia has provided insufficient assurance of the heightened rationality we have required in order to take a human life.

. . . The Court cites four reasons for shrinking from the implications of McCleskey's evidence: the desirability of discretion for actors in the criminal justice system, the existence of statutory safeguards against abuse of that discretion, the potential consequences for broader challenges to criminal sentencing, and an understanding of the contours of the judicial role. While these concerns underscore the need for sober deliberation, they do not justify rejecting evidence as convincing as McCleskey has presented.

The Court maintains that petitioner's claim "is antithetical to the fundamental role of discretion in our criminal justice system." It states that "[w]here the discretion that is fundamental to our criminal process is involved, we decline to assume that what is unexplained is invidious."

Reliance on race in imposing capital punishment, however, is antithetical to the very rationale for granting sentencing discretion. Discretion is a means, not an end. It is bestowed in order to permit the sentencer to "trea[t] each defendant in a capital case with that degree of respect due the uniqueness of the individual." The decision to impose the punishment of death must be based on a "particularized consideration of relevant aspects of the character and record of each convicted defendant." Failure to conduct such an individualized moral inquiry treats all persons convicted of a designated offense not as unique individual human beings, but as members of a faceless, undifferentiated mass to be subjected to the blind infliction of the penalty of death.

Considering the race of a defendant or victim in deciding if the death penalty should be imposed is completely at odds with this concern that an individual be evaluated as a unique human being. Decisions influenced by race rest in part on a categorical assessment of the worth of human beings according to color, insensitive to whatever qualities the individuals in question may possess. Enhanced willingness to impose the death sentence on black defendants, or diminished willingness to render such a sentence when blacks are victims, reflects a devaluation of the lives of black persons. When confronted with evidence that race more likely than not plays such a role in a capital sentencing system, it is plainly insufficient to say that the importance of discretion demands that the risk be higher before we will act—for, in such a case, the very end that discretion is designed to serve is being undermined.

Our desire for individualized moral judgments may lead us to accept some inconsistencies

in sentencing outcomes. Since such decisions are not reducible to mathematical formulae, we are willing to assume that a certain degree of variation reflects the fact that no two defendants are completely alike. There is thus a presumption that actors in the criminal justice system exercise their discretion in responsible fashion, and we do not automatically infer that sentencing patterns that do not comport with ideal rationality are suspect.

As we made clear in Batson v. Kentucky, however, that presumption is rebuttable. Batson dealt with another arena in which considerable discretion traditionally has been afforded, the exercise of peremptory challenges. Those challenges are normally exercised without any indication whatsoever of the grounds for doing so. The rationale for this deference has been a belief that the unique characteristics of particular prospective jurors may raise concern on the part of the prosecution or defense, despite the fact that counsel may not be able to articulate that concern in a manner sufficient to support exclusion for cause. As with sentencing, therefore, peremptory challenges are justified as an occasion for particularized determinations related to specific individuals, and, as with sentencing, we presume that such challenges normally are not made on the basis of a factor such as race. As we said in Batson, however, such features do not justify imposing a "crippling burden of proof," in order to rebut that presumption. The Court in this case apparently seeks to do just that. On the basis of the need for individualized decisions, it rejects evidence, drawn from the most sophisticated capital sentencing analysis ever performed, that reveals that race more likely than not infects capital sentencing decisions. The Court's position converts a rebuttable presumption into a virtually conclusive one.

The Court also declines to find McCleskey's evidence sufficient in view of "the safeguards designed to minimize racial bias in the [capital sentencing] process." Gregg v. Georgia, upheld the Georgia capital sentencing statute against a facial challenge which Justice White described in his concurring opinion as based on "simply an assertion of lack of faith" that the system could operate in a fair manner (opinion concurring in judgment). Justice White observed that the claim that prosecutors might act in an arbitrary fashion was "unsupported by any facts," and that prosecutors must be assumed to exercise their charging duties properly "[a]bsent facts to the contrary." It is clear that Gregg bestowed no permanent approval on the Georgia system. It simply held that the State's statutory safeguards were assumed sufficient to channel discretion without evidence otherwise.

It has now been over 13 years since Georgia adopted the provisions upheld in Gregg. Professor Baldus and his colleagues have compiled data on almost 2,500 homicides committed during the period 1973-1979. They have taken into account the influence of 230 nonracial variables, using a multitude of data from the State itself, and have produced striking evidence that the odds of being sentenced to death are significantly greater than average if a defendant is black or his or her victim is white. The challenge to the Georgia system is not speculative or theoretical; it is empirical. As a result, the Court cannot rely on the statutory safeguards in discounting McCleskey's evidence, for it is the very effectiveness of those safeguards that such evidence calls into question. While we may hope that a model of procedural fairness will curb the influence of race on sentencing, "we cannot simply assume that the model works as intended; we must critique its performance in terms of its results."

The Court next states that its unwillingness to regard petitioner's evidence as sufficient is based in part on the fear that recognition of McCleskey's claim would open the door to widespread challenges to all aspects of criminal sentencing. Taken on its face, such a statement seems to suggest a fear of too much justice. Yet surely the majority would acknowledge that, if striking evidence indicated that other minority groups, or women, or even persons with blond hair, were disproportionately sentenced to death, such a state of affairs would be repugnant to deeply rooted conceptions of fairness. The prospect that there may be more widespread abuse than McCleskey documents may be dismaying, but it does not justify complete abdication of our judicial role. The Constitution was framed fundamentally as a bulwark against governmental power, and preventing the arbitrary administration of punishment is a basic ideal of any society that purports to be governed by the rule of law.

In fairness, the Court's fear that McCleskey's claim is an invitation to descend a slippery slope also rests on the realization that any humanly imposed system of penalties will exhibit some imperfection. Yet to reject McCleskey's powerful evidence on this basis is to ignore both the qualitatively different character of the death penalty and the particular repugnance of racial discrimination, considerations which may properly be taken into account in determining whether various punishments are "cruel and unusual." Furthermore, it fails to take account of the unprecedented refinement and strength of the Baldus study.

It hardly needs reiteration that this Court has consistently acknowledged the uniqueness of the punishment of death. Death, in its finality, differs more from life imprisonment than a 100-year prison term differs from one of only a year or two. Because of that qualitative difference, there is a corresponding difference in the need for reliability in the determination that death is the appropriate punishment. Furthermore, the relative interests of the state and the defendant differ dramatically in the death penalty context. The marginal benefits accruing to the state from obtaining the death penalty, rather than life imprisonment, are considerably less than the marginal difference to the defendant between death and life in prison. Such a disparity is an additional reason for tolerating scant arbitrariness in capital sentencing. Even those who believe that society can impose the death penalty in a manner sufficiently rational to justify its continuation must acknowledge that the level of rationality that is considered satisfactory must be uniquely high. As a result, the degree of arbitrariness that may be adequate to render the death penalty "cruel and unusual" punishment may not be adequate to invalidate lesser penalties. What these relative degrees of arbitrariness might be in other cases need not concern us here; the point is that the majority's fear of wholesale invalidation of criminal sentences is unfounded.

The Court also maintains that accepting McCleskey's claim would pose a threat to all sentencing because of the prospect that a correlation might be demonstrated between sentencing outcomes and other personal characteristics. Again, such a view is indifferent to the considerations that enter into a determination whether punishment is "cruel and unusual." Race is a consideration whose influence is expressly constitutionally proscribed. We have expressed a moral commitment, as embodied in our fundamental law, that this specific characteristic should not be the basis for allotting burdens and benefits. Three constitutional amendments, and numerous statutes, have been prompted specifically by the desire

to address the effects of racism.

Over the years, this Court has consistently repudiated "[d]istinctions between citizens solely because of their ancestry" as being "odious to a free people whose institutions are founded upon the doctrine of equality."

Furthermore, we have explicitly acknowledged the illegitimacy of race as a consideration in capital sentencing. That a decision to impose the death penalty could be influenced by race is thus a particularly repugnant prospect, and evidence that race may play even a modest role in levying a death sentence should be enough to characterize that sentence as "cruel and unusual."

Certainly, a factor that we would regard as morally irrelevant, such as hair color, at least theoretically could be associated with sentencing results to such an extent that we would regard as arbitrary a system in which that factor played a significant role. As I have said above, however, the evaluation of evidence suggesting such a correlation must be informed not merely by statistics, but by history and experience. One could hardly contend that this Nation has, on the basis of hair color, inflicted upon persons deprivation comparable to that imposed on the basis of race. Recognition of this fact would necessarily influence the evaluation of data suggesting the influence of hair color on sentencing, and would require evidence of statistical correlation even more powerful than that presented by the Baldus study.

Furthermore, the Court's fear of the expansive ramifications of a holding for McCleskey in this case is unfounded, because it fails to recognize the uniquely sophisticated nature of the Baldus study. McCleskey presents evidence that is far and away the most refined data ever assembled on any system of punishment, data not readily replicated through casual effort. Moreover, that evidence depicts not merely arguable tendencies, but striking correlations, all the more powerful because nonracial explanations have been eliminated. Acceptance of petitioner's evidence would therefore establish a remarkably stringent standard of statistical evidence unlikely to be satisfied with any frequency.

The Court's projection of apocalyptic consequences for criminal sentencing is thus greatly exaggerated. The Court can indulge in such speculation only by ignoring its own jurisprudence demanding the highest scrutiny on issues of death and race. As a result, it fails to do justice to a claim in which both those elements are intertwined—an occasion calling for the most sensitive inquiry a court can conduct. Despite its acceptance of the validity of Warren McCleskey's evidence, the Court is willing to let his death sentence stand because it fears that we cannot successfully define a different standard for lesser punishments. This fear is baseless.

Finally, the Court justifies its rejection of McCleskey's claim by cautioning against usurpation of the legislatures' role in devising and monitoring criminal punishment. The Court is, of course, correct to emphasize the gravity of constitutional intervention, and the importance that it be sparingly employed. The fact that "[c]apital punishment is now the law in more than two thirds of our States," however, does not diminish the fact that capital punishment is the most awesome act that a State can perform. The judiciary's role in this society counts for little if the use of governmental power to extinguish life does not elicit close

scrutiny. It is true that society has a legitimate interest in punishment. Yet, as Alexander Bickel wrote:

It is a premise we deduce not merely from the fact of a written constitution but from the history of the race, and ultimately as a moral judgment of the good society, that government should serve not only what we conceive from time to time to be our immediate material needs, but also certain enduring values. This in part is what is meant by government under law.

Our commitment to these values requires fidelity to them even when there is temptation to ignore them. Such temptation is especially apt to arise in criminal matters, for those granted constitutional protection in this context are those whom society finds most menacing and opprobrious. Even less sympathetic are those we consider for the sentence of death, for execution "is a way of saying, 'You are not fit for this world, take your chance elsewhere.' " For these reasons, [t]he methods we employ in the enforcement of our criminal law have aptly been called the measures by which the quality of our civilization may be judged. Those whom we would banish from society or from the human community itself often speak in too faint a voice to be heard above society's demand for punishment. It is the particular role of courts to hear these voices, for the Constitution declares that the majoritarian chorus may not alone dictate the conditions of social life. The Court thus fulfills, rather than disrupts, the scheme of separation of powers by closely scrutinizing the imposition of the death penalty, for no decision of a society is more deserving of "sober second thought."

. . . At the time our Constitution was framed 200 years ago this year, blacks had for more than a century before been regarded as beings of an inferior order, and altogether unfit to associate with the white race, either in social or political relations; and so far inferior that they had no rights which the white man was bound to respect. Only 130 years ago, this Court relied on these observations to deny American citizenship to blacks. A mere three generations ago, this Court sanctioned racial segregation, stating that "[i]f one race be inferior to the other socially, the Constitution of the United States cannot put them upon the same plane."

In more recent times, we have sought to free ourselves from the burden of this history. Yet it has been scarcely a generation since this Court's first decision striking down racial segregation, and barely two decades since the legislative prohibition of racial discrimination in major domains of national life. These have been honorable steps, but we cannot pretend that, in three decades, we have completely escaped the grip of a historical legacy spanning centuries. Warren McCleskey's evidence confronts us with the subtle and persistent influence of the past. His message is a disturbing one to a society that has formally repudiated racism, and a frustrating one to a Nation accustomed to regarding its destiny as the product of its own will. Nonetheless, we ignore him at our peril, for we remain imprisoned by the past as long as we deny its influence in the present.

It is tempting to pretend that minorities on death row share a fate in no way connected to our own, that our treatment of them sounds no echoes beyond the chambers in which they die. Such an illusion is ultimately corrosive, for the reverberations of injustice are not so easily confined. "The destinies of the two races in this country are indissolubly linked

together," and the way in which we choose those who will die reveals the depth of moral commitment among the living.

The Court's decision today will not change what attorneys in Georgia tell other Warren McCleskeys about their chances of execution. Nothing will soften the harsh message they must convey, nor alter the prospect that race undoubtedly will continue to be a topic of discussion. McCleskey's evidence will not have obtained judicial acceptance, but that will not affect what is said on death row. However many criticisms of today's decision may be rendered, these painful conversations will serve as the most eloquent dissents of all.

# Chapter 15

# Background: Experimental Design and the Collection of Data

## 15.1 Suggested Reading on the Production of Data and Experimental Design

The Suggested Reading for this part of the book is broken down into five major areas: Replication Issues; Individual Differences; Epidemiological and Environmental Studies Generally; a collection of readings on some specific associations that have been reported in the media; and finally, a set of articles on the widely used weed killer, Atrazine, that has come under renewed and intensive scrutiny:

D.1 – Replication Issues

D.1.1 – We're So Good at Medical Studies That Most of Them Are Wrong (John Timmer, *arstechnica.com*)

D.1.2 – Keeping Computers From Ending Science's Reproducibility (John Timmer, *arstechnica.com*)

D.1.3 – Null Science: Psychology's Statistical Status Quo Draws Fire (Bruce Bower, *Science News*, June 7, 1997)

D.1.4 – Enhancing the Placebo (Olivia Judson, *New York Times*, May 3, 2010)

D.1.5 – Data From Many Drug Trials For Stroke Go Unpublished (Janet Raloff, *Science News*, April 22, 2010)

D.2 – Individual Differences

D.2.1 – When Averages Hide Individual Differences in Clinical Trials (David Kent and Rodney Hayward, *American Scientist*, January/February, 2007)

D.2.2 – Health Care: Who Knows 'Best'? (Jerome Groopman, *New York Review of Books*, February 11, 2010)

D.2.3 – In Reporting Symptoms, Don't Patients Know Best? (Denise Grady, *New York Times*, April 12, 2010)

D.2.4 – The Data-Driven Life (Gary Wolf, *New York Times*, April 26, 2010)

D.2.5 – Clues for Personalizing Breast Cancer Treatment (Tara Parker-Pope, *New York Times*, April 28, 2010)

D.2.6 – A Pregnant Question (Susan Galdos, *Science News*, June 5, 2010)

D.3 – Epidemiological and Environmental Studies Generally

D.3.1 – What if It's All Been a Big Fat Lie? (Gary Taubes, *New York Times*, July 7, 2002)

D.3.2 – Do We Really Know What Makes Us Healthy? (Gary Taubes, *New York Times*, September 16, 2007)

D.3.3 – Weighing the Evidence on Exercise (Gretchen Reynolds, *New York Times*, April 12, 2010)

D.3.4 – The Estrogen Dilemma (Cynthia Gorney, *New York Times*, April 12, 2010)

D.3.5 – Position of the American Dietetic Association: Food and Nutrition Misinformation (*Journal of the American Dietetic Association*, April 2006)

D.3.6 – That Tap Water Is Legal But May Be Unhealthy (Charles Duhigg, *New York Times*, December 17, 2009)

D.3.7 – E.P.A. to Limit Water Pollution From Mining (Tom Zeller Jr., *New York Times*, April 1, 2010)

D.3.8 – Rule on Lead Safety Set to Take Effect (Mireya Navarro, *New York Times*, April 8, 2010)

D.3.9 – Weak Rules on Toxins and Safety (David Leonhardt, *New York Times*, March 30, 2010)

D.3.10 – Skin as a Source of Drug Pollution (Janet Raloff, *Science News*, April 2, 2010)

D.3.11 – Farmers Cope With Roundup-Resistant Weeds (William Neumann and Andrew Pollack, *New York Times*, May 3, 2010)

D.3.12 – Studies Aim to Resolve Confusion Over Mercury Risks From Fish (Janet Raloff, *Science News*, April 21, 2010)

D.3.13 – Cows on Drugs (Donald Kennedy, *New York Times*, April 18, 2010)

D.3.14 – Study Says Overuse Threatens Gains From Modified Crops (Andrew Pollack, *New York Times*, April 13, 2010)

D.3.15 – The Plastic Panic (Jerome Groopman, *New Yorker*, May 31, 2010)

D.3.16 – John Rock's Error (Malcolm Gladwell, *New Yorker*, March 10, 2000)

D.4 – Some Specific Reported Associations in the Media

D.4.1 – Alcohol Distills Aggression in Large Men (Bruce Bower, *Science News*, March 5, 2010)

D.4.2 – The Spread of Superbugs (Nicholas D. Kristof, *New York Times*, March 7, 2010)

D.4.3 – Women Who Drink Gain Less Weight (Tara Parker-Pope, *New York Times*, March 8, 2010)

D.4.4 – Scientists Offer Compelling Images of Gulf War Illness (Janet Raloff, *Science News*, March 9, 2010)

D.4.5 – Drug Helps Diabetics, Trial Finds (Roni Caryn Rabin, *New York Times*, March 13, 2010)

D.4.6 – Male Infertility Linked With Prostate Cancer (Tara Parker-Pope, *New York Times*, March 23, 2010)

D.4.7 – Gulf War Syndrome Real, Institute of Medicine Concludes (Janet Raloff, *Science News*, April 10, 2010)

D.4.8 – Walnuts Slow Prostate Cancer Growth (Janet Raloff, *Science News*, March 27, 2010)

D.4.9 – Study Reports Hints of Phthalate Threat to Boys' IQs (Janet Raloff, *Science News*, April 5, 2010)

D.4.10 – FDA Says Studies on Triclosan, Used in Sanitizers and Soaps, Raise Concerns (Lyndsey Layton, *Washington Post*, April 8, 2010)

D.4.11 – Eating Chocolate Is Linked To Depression (Jennifer Corbett Booren, *Wall Street Journal*, April 27, 2010)

D.4.12 – Aging: Diet May Be Linked to Lower Alzheimer's Risk in Older People (Roni Caryn Rabin, *New York Times*, April 16, 2010)

D.4.13 – Intentional Weight Loss In Old Age Not Detrimental, Study Finds (Nathan Seppa, *Science News*, April 12, 2010)

D.4.14 – Colorectal Cancer Risk Linked To Stomach Bacterium, Inflammation (Nathan Seppa, *Science News*, April 20, 2010)

D.4.15 – Honeybee Death Mystery Deepens (Eva Everson, *Science News*, May 27, 2010)

D.4.16 – Tanning Bed Use Linked to Melanoma Risk (Nathan Seppa, *Science News*, May 27, 2010)

D.4.17 – Interphone Study Finds Hints of Brain Cancer Risk in Heavy Cell-Phone Users (Janet Raloff, *Science News*, May 17, 2010)

D.5 – The Study of Atrazine

D.5.1 – Debating How Much Weed Killer Is Safe in Your Water Glass (Charles Duhigg, *New York Times*, August 23, 2009)

D.5.2 – EPA Reviews Hints of Weed Killer's Fetal Risks (Janet Raloff, *Science News*, February 4, 2010)

D.5.3 – Weed Killer in the Crosshairs (Janet Raloff, *Science News*, February 27, 2010)

D.5.4 – Frogs: Weed Killer Creates Real Mr. Moms (Janet Raloff, *Science News*, March 1, 2010)

D.5.5 – Frogs: Clues to How Weed Killer May Feminize Males (Janet Raloff, *Science News*, March 2, 2010)

D.5.6 – Bees Face 'Unprecedented' Pesticide Exposures at Home and Afield (Janet Raloff, *Science News*, March 21, 2010)

# Chapter 16

# Ethical Considerations in Data Collection

### 16.0.1 Appendix: The Belmont Report

The Belmont Report: Ethical Principles and Guidelines for the protection of human subjects of research (April 18, 1979)

SUMMARY: On July 12, 1974, the National Research Act was signed into law, thereby creating the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. One of the charges to the Commission was to identify the basic ethical principles that should underlie the conduct of biomedical and behavioral research involving human subjects and to develop guidelines which should be followed to assure that such research is conducted in accordance with those principles. In carrying out the above, the Commission was directed to consider: (i) the boundaries between biomedical and behavioral research and the accepted and routine practice of medicine, (ii) the role of assessment of risk-benefit criteria in the determination of the appropriateness of research involving human subjects, (iii) appropriate guidelines for the selection of human subjects for participation in such research and (iv) the nature and definition of informed consent in various research settings.

The Belmont Report attempts to summarize the basic ethical principles identified by the Commission in the course of its deliberations. It is the outgrowth of an intensive four-day period of discussions that were held in February 1976 at the Smithsonian Institution's Belmont Conference Center supplemented by the monthly deliberations of the Commission that were held over a period of nearly four years. It is a statement of basic ethical principles and guidelines that should assist in resolving the ethical problems that surround the conduct of research with human subjects.

Ethical Principles & Guidelines for Research Involving Human Subjects:

Scientific research has produced substantial social benefits. It has also posed some troubling ethical questions. Public attention was drawn to these questions by reported abuses of human subjects in biomedical experiments, especially during the Second World War. During

the Nuremberg War Crime Trials, the Nuremberg code was drafted as a set of standards for judging physicians and scientists who had conducted biomedical experiments on concentration camp prisoners. This code became the prototype of many later codes intended to assure that research involving human subjects would be carried out in an ethical manner.[1]

The codes consist of rules, some general, others specific, that guide the investigators or the reviewers of research in their work. Such rules often are inadequate to cover complex situations; at times they come into conflict, and they are frequently difficult to interpret or apply. Broader ethical principles will provide a basis on which specific rules may be formulated, criticized and interpreted.

Three principles, or general prescriptive judgments, that are relevant to research involving human subjects are identified in this statement. Other principles may also be relevant. These three are comprehensive, however, and are stated at a level of generalization that should assist scientists, subjects, reviewers and interested citizens to understand the ethical issues inherent in research involving human subjects. These principles cannot always be applied so as to resolve beyond dispute particular ethical problems. The objective is to provide an analytical framework that will guide the resolution of ethical problems arising from research involving human subjects.

Part A: Boundaries Between Practice & Research

It is important to distinguish between biomedical and behavioral research, on the one hand, and the practice of accepted therapy on the other, ... to know what activities ought to undergo review for the protection of human subjects of research. The distinction between research and practice is blurred partly because both often occur together (as in research designed to evaluate a therapy) and partly because notable departures from standard practice are often called "experimental" when the terms "experimental" and "research" are not carefully defined.

For the most part, the term "practice" refers to interventions that are designed solely to enhance the well-being of an individual patient or client and that have a reasonable expectation of success. The purpose of medical or behavioral practice is to provide diagnosis, preventive treatment or therapy to particular individuals. By contrast, the term "research" designates an activity designed to test an hypothesis, permit conclusions to be drawn, and thereby to develop or contribute to generalizable knowledge (expressed, for example, in theories, principles, and statements of relationships). Research is usually described in a formal protocol that sets forth an objective and a set of procedures designed to reach that objective.[2]

[1]Since 1945, various codes for the proper and responsible conduct of human experimentation in medical research have been adopted by different organizations. The best known of these codes are the Nuremberg Code of 1947, the Helsinki Declaration of 1964 (revised in 1975), and the 1971 Guidelines (codified into Federal Regulations in 1974) issued by the U.S. Department of Health, Education, and Welfare Codes for the conduct of social and behavioral research have also been adopted, the best known being that of the American Psychological Association, published in 1973.

[2]Although practice usually involves interventions designed solely to enhance the well-being of a particular individual, interventions are sometimes applied to one individual for the enhancement of the well-being of another (e.g., blood donation, skin grafts, organ transplants) or an intervention may have the dual purpose

When a clinician departs in a significant way from standard or accepted practice, the innovation does not, in and of itself, constitute research. The fact that a procedure is "experimental," in the sense of new, untested or different, does not automatically place it in the category of research. Radically new procedures of this description should, however, be made the object of formal research at an early stage in order to determine whether they are safe and effective. Thus, it is the responsibility of medical practice committees, for example, to insist that a major innovation be incorporated into a formal research project.[3]

Research and practice may be carried on together when research is designed to evaluate the safety and efficacy of a therapy. This need not cause any confusion regarding whether or not the activity requires review; the general rule is that if there is any element of research in an activity, that activity should undergo review for the protection of human subjects.

Part B: Basic Ethical Principles

The expression "basic ethical principles" refers to those general judgments that serve as a basic justification for the many particular ethical prescriptions and evaluations of human actions. Three basic principles, among those generally accepted in our cultural tradition, are particularly relevant to the ethics of research involving human subjects: the principles of respect of persons, beneficence and justice.

1. Respect for Persons. — Respect for persons incorporates at least two ethical convictions: first, that individuals should be treated as autonomous agents, and second, that persons with diminished autonomy are entitled to protection. The principle of respect for persons thus divides into two separate moral requirements: the requirement to acknowledge autonomy and the requirement to protect those with diminished autonomy.

An autonomous person is an individual capable of deliberation about personal goals and of acting under the direction of such deliberation. To respect autonomy is to give weight to autonomous persons' considered opinions and choices while refraining from obstructing their actions unless they are clearly detrimental to others. To show lack of respect for an autonomous agent is to repudiate that person's considered judgments, to deny an individual the freedom to act on those considered judgments, or to withhold information necessary to make a considered judgment, when there are no compelling reasons to do so.

However, not every human being is capable of self-determination. The capacity for self-determination matures during an individual's life, and some individuals lose this capacity wholly or in part because of illness, mental disability, or circumstances that severely restrict

---

of enhancing the well-being of a particular individual, and, at the same time, providing some benefit to others (e.g., vaccination, which protects both the person who is vaccinated and society generally). The fact that some forms of practice have elements other than immediate benefit to the individual receiving an intervention, however, should not confuse the general distinction between research and practice. Even when a procedure applied in practice may benefit some other person, it remains an intervention designed to enhance the well-being of a particular individual or groups of individuals; thus, it is practice and need not be reviewed as research.

[3]Because the problems related to social experimentation may differ substantially from those of biomedical and behavioral research, the Commission specifically declines to make any policy determination regarding such research at this time. Rather, the Commission believes that the problem ought to be addressed by one of its successor bodies.

liberty. Respect for the immature and the incapacitated may require protecting them as they mature or while they are incapacitated.

Some persons are in need of extensive protection, even to the point of excluding them from activities which may harm them; other persons require little protection beyond making sure they undertake activities freely and with awareness of possible adverse consequence. The extent of protection afforded should depend upon the risk of harm and the likelihood of benefit. The judgment that any individual lacks autonomy should be periodically reevaluated and will vary in different situations.

In most cases of research involving human subjects, respect for persons demands that subjects enter into the research voluntarily and with adequate information. In some situations, however, application of the principle is not obvious. The involvement of prisoners as subjects of research provides an instructive example. On the one hand, it would seem that the principle of respect for persons requires that prisoners not be deprived of the opportunity to volunteer for research. On the other hand, under prison conditions they may be subtly coerced or unduly influenced to engage in research activities for which they would not otherwise volunteer. Respect for persons would then dictate that prisoners be protected. Whether to allow prisoners to "volunteer" or to "protect" them presents a dilemma. Respecting persons, in most hard cases, is often a matter of balancing competing claims urged by the principle of respect itself.

2. Beneficence. — Persons are treated in an ethical manner not only by respecting their decisions and protecting them from harm, but also by making efforts to secure their well-being. Such treatment falls under the principle of beneficence. The term "beneficence" is often understood to cover acts of kindness or charity that go beyond strict obligation. In this document, beneficence is understood in a stronger sense, as an obligation. Two general rules have been formulated as complementary expressions of beneficent actions in this sense: (1) do not harm and (2) maximize possible benefits and minimize possible harms.

The Hippocratic maxim "do no harm" has long been a fundamental principle of medical ethics. Claude Bernard extended it to the realm of research, saying that one should not injure one person regardless of the benefits that might come to others.[4] However, even avoiding harm requires learning what is harmful; and, in the process of obtaining this information, persons may be exposed to risk of harm. Further, the Hippocratic Oath requires physicians to benefit their patients "according to their best judgment." Learning what will in fact benefit may require exposing persons to risk. The problem posed by these imperatives is to decide when it is justifiable to seek certain benefits despite the risks involved, and when the benefits should be foregone because of the risks.

The obligations of beneficence affect both individual investigators and society at large, because they extend both to particular research projects and to the entire enterprise of research. In the case of particular projects, investigators and members of their institutions are obliged to give forethought to the maximization of benefits and the reduction of risk

[4]Claude Bernard (1813–1878) was a well-known French physiologist. He was the first to suggest the use of blind experiments to ensure the objectivity of scientific observation. His major work on scientific method is, *An Introduction to the Study of Experimental Medicine* (1865) ("Claude Bernard," 2011).

that might occur from the research investigation. In the case of scientific research in general, members of the larger society are obliged to recognize the longer term benefits and risks that may result from the improvement of knowledge and from the development of novel medical, psychotherapeutic, and social procedures.

The principle of beneficence often occupies a well-defined justifying role in many areas of research involving human subjects. An example is found in research involving children. Effective ways of treating childhood diseases and fostering healthy development are benefits that serve to justify research involving children—even when individual research subjects are not direct beneficiaries. Research also makes it possible to avoid the harm that may result from the application of previously accepted routine practices that on closer investigation turn out to be dangerous. But the role of the principle of beneficence is not always so unambiguous. A difficult ethical problem remains, for example, about research that presents more than minimal risk without immediate prospect of direct benefit to the children involved. Some have argued that such research is inadmissible, while others have pointed out that this limit would rule out much research promising great benefit to children in the future. Here again, as with all hard cases, the different claims covered by the principle of beneficence may come into conflict and force difficult choices.

3. Justice. — Who ought to receive the benefits of research and bear its burdens? This is a question of justice, in the sense of "fairness in distribution" or "what is deserved." An injustice occurs when some benefit to which a person is entitled is denied without good reason or when some burden is imposed unduly. Another way of conceiving the principle of justice is that equals ought to be treated equally. However, this statement requires explication. Who is equal and who is unequal? What considerations justify departure from equal distribution? Almost all commentators allow that distinctions based on experience, age, deprivation, competence, merit and position do sometimes constitute criteria justifying differential treatment for certain purposes. It is necessary, then, to explain in what respects people should be treated equally. There are several widely accepted formulations of just ways to distribute burdens and benefits. Each formulation mentions some relevant property on the basis of which burdens and benefits should be distributed. These formulations are (1) to each person an equal share, (2) to each person according to individual need, (3) to each person according to individual effort, (4) to each person according to societal contribution, and (5) to each person according to merit.

Questions of justice have long been associated with social practices such as punishment, taxation and political representation. Until recently these questions have not generally been associated with scientific research. However, they are foreshadowed even in the earliest reflections on the ethics of research involving human subjects. For example, during the 19th and early 20th centuries the burdens of serving as research subjects fell largely upon poor ward patients, while the benefits of improved medical care flowed primarily to private patients. Subsequently, the exploitation of unwilling prisoners as research subjects in Nazi concentration camps was condemned as a particularly flagrant injustice. In this country, in the 1940's, the Tuskegee syphilis study used disadvantaged, rural black men to study the untreated course of a disease that is by no means confined to that population. These subjects

were deprived of demonstrably effective treatment in order not to interrupt the project, long after such treatment became generally available.

Against this historical background, it can be seen how conceptions of justice are relevant to research involving human subjects. For example, the selection of research subjects needs to be scrutinized in order to determine whether some classes (e.g., welfare patients, particular racial and ethnic minorities, or persons confined to institutions) are being systematically selected simply because of their easy availability, their compromised position, or their manipulability, rather than for reasons directly related to the problem being studied. Finally, whenever research supported by public funds leads to the development of therapeutic devices and procedures, justice demands both that these not provide advantages only to those who can afford them and that such research should not unduly involve persons from groups unlikely to be among the beneficiaries of subsequent applications of the research.

Part C: Applications

Applications of the general principles to the conduct of research leads to consideration of the following requirements: informed consent, risk/benefit assessment, and the selection of subjects of research.

1. Informed Consent: Respect for persons requires that subjects, to the degree that they are capable, be given the opportunity to choose what shall or shall not happen to them. This opportunity is provided when adequate standards for informed consent are satisfied.

While the importance of informed consent is unquestioned, controversy prevails over the nature and possibility of an informed consent. Nonetheless, there is widespread agreement that the consent process can be analyzed as containing three elements: information, comprehension and voluntariness.

Information: Most codes of research establish specific items for disclosure intended to assure that subjects are given sufficient information. These items generally include: the research procedure, their purposes, risks and anticipated benefits, alternative procedures (where therapy is involved), and a statement offering the subject the opportunity to ask questions and to withdraw at any time from the research. Additional items have been proposed, including how subjects are selected, the person responsible for the research, etc.

However, a simple listing of items does not answer the question of what the standard should be for judging how much and what sort of information should be provided. One standard frequently invoked in medical practice, namely the information commonly provided by practitioners in the field or in the locale, is inadequate since research takes place precisely when a common understanding does not exist. Another standard, currently popular in malpractice law, requires the practitioner to reveal the information that reasonable persons would wish to know in order to make a decision regarding their care. This, too, seems insufficient since the research subject, being in essence a volunteer, may wish to know considerably more about risks gratuitously undertaken than do patients who deliver themselves into the hand of a clinician for needed care. It may be that a standard of "the reasonable volunteer" should be proposed: the extent and nature of information should be such that persons, knowing that the procedure is neither necessary for their care nor perhaps fully understood, can decide whether they wish to participate in the furthering of knowledge. Even

when some direct benefit to them is anticipated, the subjects should understand clearly the range of risk and the voluntary nature of participation.

A special problem of consent arises where informing subjects of some pertinent aspect of the research is likely to impair the validity of the research. In many cases, it is sufficient to indicate to subjects that they are being invited to participate in research of which some features will not be revealed until the research is concluded. In all cases of research involving incomplete disclosure, such research is justified only if it is clear that (1) incomplete disclosure is truly necessary to accomplish the goals of the research, (2) there are no undisclosed risks to subjects that are more than minimal, and (3) there is an adequate plan for debriefing subjects, when appropriate, and for dissemination of research results to them. Information about risks should never be withheld for the purpose of eliciting the cooperation of subjects, and truthful answers should always be given to direct questions about the research. Care should be taken to distinguish cases in which disclosure would destroy or invalidate the research from cases in which disclosure would simply inconvenience the investigator.

Comprehension: The manner and context in which information is conveyed is as important as the information itself. For example, presenting information in a disorganized and rapid fashion, allowing too little time for consideration or curtailing opportunities for questioning, all may adversely affect a subject's ability to make an informed choice.

Because the subject's ability to understand is a function of intelligence, rationality, maturity and language, it is necessary to adapt the presentation of the information to the subject's capacities. Investigators are responsible for ascertaining that the subject has comprehended the information. While there is always an obligation to ascertain that the information about risk to subjects is complete and adequately comprehended, when the risks are more serious, that obligation increases. On occasion, it may be suitable to give some oral or written tests of comprehension.

Special provision may need to be made when comprehension is severely limited—for example, by conditions of immaturity or mental disability. Each class of subjects that one might consider as incompetent (e.g., infants and young children, mentally disable patients, the terminally ill and the comatose) should be considered on its own terms. Even for these persons, however, respect requires giving them the opportunity to choose to the extent they are able, whether or not to participate in research. The objections of these subjects to involvement should be honored, unless the research entails providing them a therapy unavailable elsewhere. Respect for persons also requires seeking the permission of other parties in order to protect the subjects from harm. Such persons are thus respected both by acknowledging their own wishes and by the use of third parties to protect them from harm.

The third parties chosen should be those who are most likely to understand the incompetent subject's situation and to act in that person's best interest. The person authorized to act on behalf of the subject should be given an opportunity to observe the research as it proceeds in order to be able to withdraw the subject from the research, if such action appears in the subject's best interest.

Voluntariness: An agreement to participate in research constitutes a valid consent only if voluntarily given. This element of informed consent requires conditions free of coercion and

undue influence. Coercion occurs when an overt threat of harm is intentionally presented by one person to another in order to obtain compliance. Undue influence, by contrast, occurs through an offer of an excessive, unwarranted, inappropriate or improper reward or other overture in order to obtain compliance. Also, inducements that would ordinarily be acceptable may become undue influences if the subject is especially vulnerable.

Unjustifiable pressures usually occur when persons in positions of authority or commanding influence—especially where possible sanctions are involved—urge a course of action for a subject. A continuum of such influencing factors exists, however, and it is impossible to state precisely where justifiable persuasion ends and undue influence begins. But undue influence would include actions such as manipulating a person's choice through the controlling influence of a close relative and threatening to withdraw health services to which an individual would otherwise be entitled.

2. Assessment of Risks and Benefits: The assessment of risks and benefits requires a careful arrayal of relevant data, including, in some cases, alternative ways of obtaining the benefits sought in the research. Thus, the assessment presents both an opportunity and a responsibility to gather systematic and comprehensive information about proposed research. For the investigator, it is a means to examine whether the proposed research is properly designed. For a review committee, it is a method for determining whether the risks that will be presented to subjects are justified. For prospective subjects, the assessment will assist the determination whether or not to participate.

The Nature and Scope of Risks and Benefits: The requirement that research be justified on the basis of a favorable risk/benefit assessment bears a close relation to the principle of beneficence, just as the moral requirement that informed consent be obtained is derived primarily from the principle of respect for persons. The term "risk" refers to a possibility that harm may occur. However, when expressions such as "small risk" or "high risk" are used, they usually refer (often ambiguously) both to the chance (probability) of experiencing a harm and the severity (magnitude) of the envisioned harm.

The term "benefit" is used in the research context to refer to something of positive value related to health or welfare. Unlike, "risk," "benefit" is not a term that expresses probabilities. Risk is properly contrasted to probability of benefits, and benefits are properly contrasted with harms rather than risks of harm. Accordingly, so-called risk/benefit assessments are concerned with the probabilities and magnitudes of possible harm and anticipated benefits. Many kinds of possible harms and benefits need to be taken into account. There are, for example, risks of psychological harm, physical harm, legal harm, social harm and economic harm and the corresponding benefits. While the most likely types of harms to research subjects are those of psychological or physical pain or injury, other possible kinds should not be overlooked.

Risks and benefits of research may affect the individual subjects, the families of the individual subjects, and society at large (or special groups of subjects in society). Previous codes and Federal regulations have required that risks to subjects be outweighed by the sum of both the anticipated benefit to the subject, if any, and the anticipated benefit to society in the form of knowledge to be gained from the research. In balancing these different

elements, the risks and benefits affecting the immediate research subject will normally carry special weight. On the other hand, interests other than those of the subject may on some occasions be sufficient by themselves to justify the risks involved in the research, so long as the subjects' rights have been protected. Beneficence thus requires that we protect against risk of harm to subjects and also that we be concerned about the loss of the substantial benefits that might be gained from research.

The Systematic Assessment of Risks and Benefits: It is commonly said that benefits and risks must be "balanced" and shown to be "in a favorable ratio." The metaphorical character of these terms draws attention to the difficulty of making precise judgments. Only on rare occasions will quantitative techniques be available for the scrutiny of research protocols. However, the idea of systematic, nonarbitrary analysis of risks and benefits should be emulated insofar as possible. This ideal requires those making decisions about the justifiability of research to be thorough in the accumulation and assessment of information about all aspects of the research, and to consider alternatives systematically. This procedure renders the assessment of research more rigorous and precise, while making communication between review board members and investigators less subject to misinterpretation, misinformation and conflicting judgments. Thus, there should first be a determination of the validity of the presuppositions of the research; then the nature, probability and magnitude of risk should be distinguished with as much clarity as possible. The method of ascertaining risks should be explicit, especially where there is no alternative to the use of such vague categories as small or slight risk. It should also be determined whether an investigator's estimates of the probability of harm or benefits are reasonable, as judged by known facts or other available studies.

Finally, assessment of the justifiability of research should reflect at least the following considerations: (i) Brutal or inhumane treatment of human subjects is never morally justified. (ii) Risks should be reduced to those necessary to achieve the research objective. It should be determined whether it is in fact necessary to use human subjects at all. Risk can perhaps never be entirely eliminated, but it can often be reduced by careful attention to alternative procedures. (iii) When research involves significant risk of serious impairment, review committees should be extraordinarily insistent on the justification of the risk (looking usually to the likelihood of benefit to the subject—or, in some rare cases, to the manifest voluntariness of the participation). (iv) When vulnerable populations are involved in research, the appropriateness of involving them should itself be demonstrated. A number of variables go into such judgments, including the nature and degree of risk, the condition of the particular population involved, and the nature and level of the anticipated benefits. (v) Relevant risks and benefits must be thoroughly arrayed in documents and procedures used in the informed consent process.

3. Selection of Subjects: Just as the principle of respect for persons finds expression in the requirements for consent, and the principle of beneficence in risk/benefit assessment, the principle of justice gives rise to moral requirements that there be fair procedures and outcomes in the selection of research subjects.

Justice is relevant to the selection of subjects of research at two levels: the social and

the individual. Individual justice in the selection of subjects would require that researchers exhibit fairness: thus, they should not offer potentially beneficial research only to some patients who are in their favor or select only "undesirable" persons for risky research. Social justice requires that distinction be drawn between classes of subjects that ought, and ought not, to participate in any particular kind of research, based on the ability of members of that class to bear burdens and on the appropriateness of placing further burdens on already burdened persons. Thus, it can be considered a matter of social justice that there is an order of preference in the selection of classes of subjects (e.g., adults before children) and that some classes of potential subjects (e.g., the institutionalized mentally infirm or prisoners) may be involved as research subjects, if at all, only on certain conditions.

Injustice may appear in the selection of subjects, even if individual subjects are selected fairly by investigators and treated fairly in the course of research. Thus injustice arises from social, racial, sexual and cultural biases institutionalized in society. Thus, even if individual researchers are treating their research subjects fairly, and even if IRBs are taking care to assure that subjects are selected fairly within a particular institution, unjust social patterns may nevertheless appear in the overall distribution of the burdens and benefits of research. Although individual institutions or investigators may not be able to resolve a problem that is pervasive in their social setting, they can consider distributive justice in selecting research subjects.

Some populations, especially institutionalized ones, are already burdened in many ways by their infirmities and environments. When research is proposed that involves risks and does not include a therapeutic component, other less burdened classes of persons should be called upon first to accept these risks of research, except where the research is directly related to the specific conditions of the class involved. Also, even though public funds for research may often flow in the same directions as public funds for health care, it seems unfair that populations dependent on public health care constitute a pool of preferred research subjects if more advantaged populations are likely to be the recipients of the benefits.

One special instance of injustice results from the involvement of vulnerable subjects. Certain groups, such as racial minorities, the economically disadvantaged, the very sick, and the institutionalized may continually be sought as research subjects, owing to their ready availability in settings where research is conducted. Given their dependent status and their frequently compromised capacity for free consent, they should be protected against the danger of being involved in research solely for administrative convenience, or because they are easy to manipulate as a result of their illness or socioeconomic condition.

# Chapter 17

# The Federal Rules of Evidence

## 17.1 Suggested Reading on the Admissibility of Expert Testimony, The Federal Rules of Evidence, and Related Topics

D.6 – Legal Admissibility of Expert Testimony

D.6.1 – Daubert v. Merrell Dow Pharmaceuticals, Inc.: The Battle Over Admissibility Standards for Scientific Evidence in Court (Suzanne Orofino, *Journal of Undergraduate Studies*, *3*, 109–111, Summer, 1996)

D.6.2 – Science and Society: The Interdependence of Science and Law (Stephen Breyer, *Science*, April 24, 1998)

D.6.3 – Something Rotten At the Core of Science? (David F. Horrobin, *Trends in Pharmacological Sciences*, February, 2001)

D.6.4 – Law Revises Standards for Scientific Study (Andrew C. Revkin, *New York Times*, March 21, 2002)

D.6.5 – 'Data Quality' Law Is Nemesis Of Regulation (Rick Weiss, *The Washington Post*, August 16, 2004)

D.6.6 – Is Science Different for Lawyers? (David L. Faigman, *Science*, July 19, 2002)

D.6.7 – Scientific Evidence and Public Policy (David Michaels, *American Journal of Public Health*, Supplement 1, 2005)

D.6.8 – Regulatory Parallels to Daubert: Stakeholder Influence, "Sound Science," and the Delayed Adoption of Health-Protective Standards (Roni A. Neff and Lynn R. Goldman, *American Journal of Public Health*, Supplement 1, 2005)

D.6.9 – Doubt Is Their Product (David Michaels, *Scientific American*, June, 2005)

D.6.10 – Putting Science in the Dock (Barry Yeoman, *The Nation*, March 26, 2007)

D.6.11 – Science Sooner: Moving Daubert to an Earlier Stage of Litigation (David Cooner and Sane Riester, *Daubert On-line*, May 22, 2009)

# Chapter 18

# Some Concluding Remarks